Studies in Public Choice

Joshua Hall · Marcus Witcher *Editors*

Public Choice Analyses of American Economic History

Volume 1

Springer

Studies in Public Choice

Volume 35

More information about this series at http://www.springer.com/series/6550

Joshua Hall • Marcus Witcher
Editors

Public Choice Analyses of American Economic History

Volume 1

Springer

Editors
Joshua Hall
Department of Economics
West Virginia University
Morgantown, West Virginia, USA

Marcus Witcher
West Virginia University
Morgantown, West Virginia, USA

ISSN 0924-4700
Studies in Public Choice
ISBN 978-3-319-77591-3 ISBN 978-3-319-77592-0 (eBook)
https://doi.org/10.1007/978-3-319-77592-0

Library of Congress Control Number: 2018938802

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Buchanan and Tullock and Downs
and Riker and too many others to mention.
We stand on your shoulders.

Acknowledgements

We would like to thank the Center for Free Enterprise at West Virginia University for support for this project. Collin Hodges provided invaluable research assistance.

Contents

Contributors

Jeremy Atack Vanderbilt University, Nashville, TN, USA

Fred Bateman (deceased 1937–2012) University of Georgia, Athens, GA, USA

Dora L. Costa University of California, Los Angeles, CA, USA

Vincent Geloso Texas Tech University, Lubbock, TX, USA

Jac C. Heckelman Wake Forest University, Winston-Salem, NC, USA

Phillip W. Magness Berry College, Mount Berry, GA, USA

Robert A. McGuire The University of Akron, Akron, OH, USA

David Mitchell University of Central Arkansas, Conway, AR, USA

John Joseph Wallis University of Maryland, College Park, MD, USA

John H. Wood Wake Forest University, Winston-Salem, NC, USA

Chapter 1
British Public Debt, the Acadian Expulsion and the American Revolution

Vincent Geloso

Abstract Starting in 1755, the French-speaking colonists of Atlantic Canada (known as the Acadians) were deported by the British. The expulsion was desired by the American colonists in New England but was ultimately opposed by the British government. In fact, the expulsion was enacted against the wishes of the Imperial government. Set against the backdrop of rising public debt in Britain, the costly expulsion of the Acadians (combined with the subsequent conquest of the French-speaking colony of Quebec) contributed to a change in policy course favoring centralization. Using public choice theory, I construct a narrative to argue that the Acadian expulsion contributed to the initiation of the American Revolution.

1.1 Introduction

Until the end of the French and Indian War (1754–1763), the American colonies faced a more or less lenient British Crown—an era that Edmund Burke qualified as one of salutary neglect for the Americans. Throughout the period, the British Parliament and the Crown were lax in the enforcement of laws and edicts in the colonies. The colonies were given a certain degree of autonomy that differed from those of other colonies in the Americas (Bordo and Cortés-Conde 2001). They were also lightly taxed given the level of British public spending in the colonies (Davis and Huttenback 1982). However, a shift in policy began to emerge during with the onset of the French and Indian War and it picked up steam at its end. The shift related to requiring a greater financial contribution to the Empire from the American colonists.

What explains this shift away from salutary neglect? The most accepted narrative is that the debt-and-tax wary British felt that they were paying for the heavy burden of the Empire while the American colonists were in effect benefiting from

V. Geloso (✉)
Texas Tech University, Lubbock, TX, USA
e-mail: vincent.geloso@ttu.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_1

the Empire. The burden of the public debt was the dominating political issue in Britain at the time and it formed the backdrop of most policy debates. In fact, the issue of the public debt in Britain was key to the supremacy of the Whigs in British politics (Stasavage 2007, 2003). Asking the colonies to contribute to the financing of the Empire was seen in Britain as a reasonable policy given the state of public finances (Rabushka 2008, 730–732). Simultaneously, the American colonists claimed that they ought to obtain more autonomy with regards to their governance (de Figueiredo Jr et al. 2006, 391–394). The demands of both parties being impossible to reconcile, political separation seemed inevitable.

However, the attempt to tax Americans was complemented by an attempt by colonial officials in England—through the Board of Trade—to become more active in the management of the colony (Geloso 2015; Greene 1986, 1994; Speck 1994). The policy shift was one towards a more direct involvement on the part of Britain in the management of the colony. Governors would be given extensive sets of guidelines to follow with very detailed policies to enact. Privileges previously enjoyed by the colonies were increasingly challenged by the Board of Trade and the Crown. This constituted an important contributor to the initiation of the American Revolution.

It is this component of the shift away from salutary neglect that is the concern of the present paper. The shift is analyzed through the lenses of public choice theory and the role played by two substantial events that motivated the shift: the Acadian Upheaval of the 1750s and the Conquest of Quebec. This paper will argue that the policy of salutary neglect had, by the 1750s, given the colonies virtual self-governance with respect to their internal affairs (de Figueiredo Jr et al. 2006, 393). However, this autonomy did not come with full fiscal responsibility as most of the financial costs of some crucial political decisions ended up being shouldered by Britain. Colonial governments were responding to the demands of settlers who wanted to expand settlement westwards and, in the case of New England, towards the northeast as well. In that latter case, the colonists saw their way barred by the French-Canadians settlers known as the Acadians in the modern-day Canadian provinces of New Brunswick, Prince Edward Island and Nova Scotia. The Acadians, who were in a strange legal situation that led them to be neither loyal to France or England, entertained some of the most friendly relations with Native Indians (most notably the Mi'kmaq). These two groups acted as a barrier to the settlement of New England farmers who lobbied colonial officials to deport the Acadians and squash the Indian tribe. Eventually, they got their way and the majority of the Acadian population was deported against the explicit wishes and orders from the Crown not to proceed with the deportation plan. A costly endeavor for such a minor theatre of operation, the financial cost of the Acadian Upheaval was shouldered largely by Britain. The fact that local politicians in America could take decisions whose burden would be shouldered elsewhere is a classic case of rent-seeking. This problem also incentivized the Crown, Parliament, and the Board of Trade to favor centralization in order to avoid a repetition of the problem once the more heavily (French) populated colony of Quebec was ceded by France to Britain.

This paper is organized in three sections. The first section briefly overviews the background of British public finances to properly put into perspective the importance of the Acadian Upheaval. The second section discusses the Acadian Upheaval and its costs. The last section explains, through the lenses of elements drawn from public choice theory, how the elements described in the first two sections contributed to the policy shift.

1.2 British Public Finance and the Burden of Empire

For most of the eighteenth century, England was locked in intermittent conflict with France and its varying allies. This long century of war meant that the British state had to grow extensively in order to mobilize the resources required for war (O'Brien 1988, 2011) and this meant a rapidly growing public debt. In the last year of the single longest period of peace in Europe, from 1713 to 1740, the public debt of Britain stood at 75% of the economy (Clark 2001, 435). By the end of the French and Indian War, it had surged to 132% and it fluctuated close to the 100% line until the American Revolution (see panel 3 of Table 1.1 for averages over periods of peace and war). As such, the debt constituted the premier political issue in Britain, especially since ownership of government securities became more widespread throughout the era. It was also an important factor in insuring a bargain between different factions that coalesced to form the Whig party which had a supremacy in government from 1714 to 1760.

This latter point is crucial to understand the rationale of political actors in Britain. Within the Whig party, there was a wide array of differing actors which included government creditors who bargained with religious dissenters and aristocrats who sought an expansion of parliamentary powers. Its main opponent was the Tory Party which had the landed aristocracy as one of its main constituents. That constituency

Table 1.1 British government spending in America, public debt and size of the British government

Years	Spending in America as share of total expenditures (%)	Spending in America as share of military expenditures (Navy+Army) (%)	British public debt as a share of GDP (%)	British government spending as share of the economy (%)
1740–1748	1.6	2.7	83.6	9.8
1749–1755	3.6	8.2	97.1	7.5
1756–1763	6.7	10.3	98.8	13.3
1764–1775	4.1	11.0	117.4	8.6

Sources: For British government spending in America: Gwyn (1980, 77); source for British expenditures: Mitchell (1988, 579); source for public debt: Clark (2001, 435); source for size of the economy: Broadberry et al. (2015)

generally chafed under the burden of the debt since the lion's share of taxation efforts fell upon it. The presence of government creditors within the Whig coalition led to the design of policies and constraints meant to prevent debt default and insure repayment. Moreover, these creditors did derive earnings from these securities whose repudiation risks they were trying to minimize. As such, the coalition held together as it benefitted this group (Stasavage 2003). Combined with the generous patronage conferred under the long ministry of Robert Walpole (1721–1742), the coalition was able to maintain itself in power and push the Tories into opposition for many decades (Eccleshall and Walker 1998, 2). The long-run legacy of this power arrangement was that debt-ownership became more widespread. This meant that debt creditors as a political base of support only expanded.[1] As such, policies that insured the financial credibility of the British government were at the center of the rationale of political actors during the eighteenth century.

With this perspective in mind, the growing cost of the Empire's reach into North America takes its relevance. Most of the financial resources allocated to North America were military in nature (Gwyn 1980, 77) and, as a share of the total military budget of Britain, they kept increasing: from less than 3% of the military budget in 1740 to above 11% by 1775 (see Table 1.1). The view in Britain was that the American colonists needed to pay their share of the Empire's defense burden. There was cause in this argument since it seems that even after the tax hikes of the 1763–1775 era, the colonies were still net recipients of British government spending (Davis and Huttenback 1982). Self-interested political actors viewed, within the overarching framework of the public debt, positively any policy that would control the net cost of the colonies.

1.3 The Acadian Upheaval

The combination of the relevance of the public debt issue in British politics and the rising costs of protecting the American colonies explains the role that the Acadian Upheaval played in the policy shift towards centralization.

In 1713, the population of French descent that populated today's Maritime provinces of Canada (but mostly Nova Scotia) fell under British rule. However, a strange political bargain (imposed by the fact that the population was Catholic) was struck whereby the population (known as the Acadians) did not have to take a loyalty oath to the King. Rather they had to take an oath of neutrality (Faragher 2005, 125–150). Few Acadians took the oath. In any case, they were largely indifferent to British rule (Akins 1869, 11) which made the British distrustful. The British

[1]It is worth pointing out that William Pitt the Younger, a *New Tory*, was actually well-versed in financial matters. During his premierships (1783–1801, 1804–1806) that overlapped with the French Wars (1792–1815) when public debt expanded rapidly, Pitt developed a sinking fund to credibly commit to the repayment of the public debt. This shift within those of Tory sympathies shows how much the landscape of British politics had changed.

settlers in neighboring New England were especially distrustful. The issue was compounded by the fact that the Acadians entertained very friendly economic and cultural relations with the Native Indians of Mi'kmaq tribes who entertained less cordial relations with the American settlers of New England (Faragher 2005, 179).

The combination of these facts meant that there was room to exploit the distrustful disposition of the British who entertained ideas regarding deportation (Akins 1869, 6,9,58–59,69–70). Settlers from neighboring New England and British-held Nova Scotia used the discomfort surrounding the neutrality oath to further their interests. First, the growth of the Acadian population was faster than theirs, which they perceived as potential competition for land (Johnston 2003, 41–42). Second, the Acadians were known for developing, impressively given the capital constraints associated with frontier economies such as those of the New World, elaborate irrigation systems to permit agricultural growth. The deportation of the Acadians, under the pretext of their constituting a credible security threat, would have meant the ability for American colonists to settle already improved lands (Akins 1869, 9). Third, the friendly disposition of the Acadians towards the Mi'kmaq who had much less cordial relations with the American settlers meant that the strength of the tribe would diminish if the Acadians were deported.[2] These motivations were sufficient enough for British settlers to push for deportation arguing that on top of being a threat, the Acadians were "lazy [and] unskillful in methods of agricultural [and cannot] be led or drove into a better way of thinking" (Akins 1869, 102) and that "industrious laborers" of British stock would be able to farm the land (Akins 1869, 9).

The drive for deportation was largely made by settlers in New England and British Nova Scotia. This was resisted by the government in London. In the 1750s, the Secretary of State underlined "pernicious consequences" from the deportation of an otherwise "considerable number of useful subjects" (Akins 1869, 581–582). When the deportation began on August 10th 1755, London had not sanctioned the move. In fact, London had sent instructions not to proceed with the deportation on August 13th 1755. The Imperial government had, in effect, rejected the move but it ended up with the bill. Overall, close to 7000 individuals were deported and the military and transport costs of the operation were estimated at somewhere between 87,621£ and 101,886£ (see Table 1.2) (Geloso 2015, 58–63). For an operation of arguably minor military importance, the costs were considerable: somewhere between 1.2% and 1.4% of British government spending and between 2.7% and 3.2% of total military expenditures. It also represented between 32.7% and 38% of the average annual budget dedicated to North America. Given the minor importance of the colony and the lack of a perceived threat on the part of London, these constituted heavy costs for the Crown.

[2]Early on, the British governors of Massachusetts offered payments for the scalps of any Mi'kmaq (regardless of age and gender) (Akins 1869, 581–589).

Table 1.2 The costs of the deportation of the Acadians

	High estimate	Low estimate
Total cost	101,886£	87,621£
Cost per deportee	14.56£	12.52£
As share of British government spending	1.4%	1.2%
As share of British expenditures on N America	38.0%	32.7%

Source: Costs: Geloso (2015, 62); source for spending in Britain and in North America: see Table 1.1

Note: Geloso, presented the cost per head as the cost relative to the total Acadian population which was in the logic of his article. Here, it is more logical to show the cost per deportee. However, Plank (2003, 149) proposes a much higher figure for deportation which nears 11,000 (the near totality of the population). Most historians agree that the number of deportees is between 7000 and 8000. There were also supply costs and recolonization expenses that were not included in these figures which amount to 14,570£ and 62,655£ (Geloso 2015, 62)

1.4 A Public Choice Interpretation of the Policy Shift

The fact that the Imperial government had explicitly rejected the project while footing the bill is crucial to our understanding of the shift towards increased centralization. Indeed, the colonists obtained the benefits of the deportation—available land combined with a parliamentary grant for recolonization—but they did not assume the costs. The level of autonomy conceded to the American colonies did not come with the financial responsibility of autonomy and this institutional arrangement formed a "moral hazard" problem for the Empire (Geloso 2015, 72). This problem is well explained by a public choice interpretation that relies on the concept of rent-seeking. It is from that explanation that the policy shift can be understood properly.

The settlers of New England and British Nova Scotia were engaging in rent-seeking (Tullock 1967; Krueger 1974). Their efforts to link the prejudices running against French-speaking Catholic with a potential security concern were deployed in order to redistribute land in their favor. In fact, they were quite explicit about it. Faragher (2006, 83) points to a letter (echoing many others) published in newspapers of New York, Pennsylvania and Maryland that presented the expulsion of the Acadians as a "Great and Noble Scheme" and as a chance to obtain land "as good (...,) as any in the world".

However, the actors who took the decision to expel the Acadians did not have to assume the full burden of the decision. On the first hand, governors William Shirley of Massachusetts and Charles Lawrence of Nova Scotia were largely respondent to the colonists who constantly lobbied for the expulsion. Moreover, the "salutary neglect" of Britain meant that they had a free hand in responding to this constituency. On the other hand, this autonomy did not come with full fiscal responsibility. While, the colonies did collect taxes for their own civil governments (Rabushka 2008) and militia service represented their defense contribution, the regular troops and the ships of the Royal Navy were provided by Britain. The latter

constituted the lion's share of colonial expenses. Local militiamen, untrained and personally assuming the expenses of their equipment, on leave from their farms or trades could have hardly afforded to spend a season campaigning in order to expel a few thousand settlers. Nor would they have had the capacity. The expulsion required the use of 2250 soldiers who had to scout for stragglers and repel the Acadians who evaded capture and fought back, with the help of the Mi'kmaq, against them (Geloso 2015, 60–61). Moreover, frigates from the squadron assigned to North America waters had to be used to protect the large number of transports ships.

Finally, it is worth pointing out that the Acadians and the Native Indians had developed an understanding that new settlers would have had to respect. The Acadians ceded the wooded uplands to the Mi'kmaq for migration, fishing, hunting and gathering. They confined their settlements to the coastal lowlands (where they built elaborate dykes to permit farming) and both parties would exchange furs for agricultural produces (Faragher 2005, 2006). In the framework of the less cordial relationship between British colonists and the Native Indians, raids and conflicts were more frequent than trade. Absent the implicit subsidy on the part of Britain, they would have had to assume the extra burden of fighting the natives. The fact that taxpayers in Britain shouldered the largest share of the costs of the operation (and very little of the benefits given the strategic relevance of expelling the Acadians) meant that the colonists did not have the fiscal responsibility that came with their great autonomy.

This situation corresponds to a problem of dispersed costs and concentrated benefits. This type of problem emerges when choices are divorced from the consequences of those choices (i.e. a lack of responsibility). The sectional interests of Massachusetts and Nova Scotia competed for policies whose benefits would accrue largely to them. However, they would have been less likely to pursue those benefits had the costs not been assumed by a third party. Such a problem would be consistent with the observation that the colonists were net recipients of government spending from Britain (Davis and Huttenback 1982).

These problems were not lost on political actors in Britain. Following the Acadian Upheaval and the French and Indian War, parliamentary committees increased their scrutiny of the public expenses of the colony of Nova Scotia and credits were withheld from the colony if it did not get its accounts in order (Hully 2012, 122–123). In Parliament, Edmund Burke lambasted the British colony of Nova Scotia. He and his followers considered the colony of Halifax as the "greatest public pork barrel project yet opened in North America" (Griffiths 1992, 79). Indeed, the Crown had expanded—over the course of the colony's existence—close to 700,000£ (Gwyn 1998, 28). The Board of Trade felt compelled—after being submitted to political pressures—to exercise tight control over the colonies (i.e. greater centralization), starting with Nova Scotia. The strict control that the Board of Trade exercised over Nova Scotia (i.e. greater centralization) formed a sort of training ground for further attempts at centralization (Hully 2012). The impetus for centralization was further increased when Britain conquered the colony of Quebec (populated by 60,000 Catholic inhabitants of French descent). The Board of Trade and the Secretary of State issued very detailed directives to the governors

of Quebec—82 sets of instructions were delivered to Governor James Murray which dealt with topics from marriage, relations with the Indians, iron works and religion (Shortt and Doughty 1918, 132–149).

However, centralization did not entail "one-size-fits-all". In fact, it was seen as a tool for flexibility and better coordination (Lawson 1989) where policies could custom-made for each colony. From this flexibility emerged an informal policy of toleration[3] towards the French-speaking Catholic population of Quebec which culminated, by 1774, in the Quebec Act which recognized many of the civil institutions that had existed prior to the conquest. This limited the risks of rebellions[4] and alliances with Native Indians[5] while also preserving the revenues associated with the fur trade (the main export from Quebec). These reduced risks allowed the British to dramatically wind down the garrison from 3200 men (for Quebec City only) to less than 1000 (Geloso 2015, 71) between 1763 and 1771. It also allowed the Governors of Quebec to face down the merchants who wanted tough measures adopted against the French-Canadians and more legislative autonomy (but only for the very few Protestants, not the Catholic majority) (Lawson 1989).

Toleration extended in a different form to the Native Indians. Wars with them had been prohibitively costly. Policies meant to appease and soothe the relations with them were seen in an increasingly favorable eye. If westwards settlement could be limited or be accomplished in a more agreeable manner between the involved parties, this would limit the probability of conflict. Moreover, bribing Indians meant creating a pool of loyal allies who could complement British military forces. At the end of the French and Indian War, such approaches were adopted in all earnestness by the British to the point that historian Robert Allen (1992) could title his book on British Indian policy and the defense of Canada as *His Majesty's Indian Allies*.

Toleration was a cost-reducing policy tailored specifically for Quebec and its adoption was heavily informed by the Acadian Upheaval (Geloso 2015).[6] Increased

[3]Here, I do not mean toleration in the enlightenment sense, even less in the sense of acceptation. Rather, I mean a form of pragmatic shrugging-off of differences in order to serve other purposes. Indeed, many of the government officials of Quebec, with the strange exception of Governor James Murray, held negative views of the French-Canadians. These men (which included Governor Guy Carleton who replaced Murray) nonetheless pushed for toleration of the civil and religious institutions of the French-Canadians.

[4]The policy paid off early on as many Canadians refused to help the Americans during the 1775 invasion—some did, but many more fought for the British and a great deal more simply engaged in a friendly neutrality towards the British.

[5]The policy seems to have paid off in that regard as well. During the Pontiac uprising of the early 1760s, French-Canadian militiamen heeded the call of the governor to help them fight the Native Indians. Although they were not actively engaged, their willingness to fight on the behalf of the British suggests that the toleration policy paid off.

[6]There is supportive evidence that this bred further waves of toleration as government policy. Indeed, it is argued that the Quebec Act of 1774—which formalized the toleration policy—formed the basis of the Irish Catholic Relief Act of 1778 which eased many constraints placed upon Catholics in Ireland (Stanbridge 2003).

centralization permitted the adoption of policies tailored to each individual colony. The policy response lends itself to public choice analysis. Readers should recall that, as mentioned above, public debt ownership in Britain became more widespread throughout the eighteenth century. Any policy courses affecting the public debt would resonate in the mind of the (propertied male) voters of Britain. Moreover, an increasing number of government creditors were part of the governing apparatus. In the functions they occupied, they pushed for the adoption of policies that insured a credible commitment towards debt repayment (i.e. reducing default risks). The fact that they shouldered the costs of policies that benefitted only the American colonists would have made these entrenched interests more likely to respond to costly endeavors like the Acadian Upheaval. As politicians with vested interests, they were acting on their behalf even if it meant a tighter control of expenditures than would have otherwise been the case. The policy response was thus a rent-seeking response to a problem created by earlier rent-seeking.

This has an important implication for American history. The attempt to centralize on the part of London was a crucial ingredient in the initiation of hostilities between the colonists and Britain. The Acadian fiasco was a key ingredient in that policy shift. As such, it contributed in no small part to the American Revolution. The Acadian Upheaval also informed Britain of the need to adopt a pragmatic policy of toleration once Quebec was acquired which could be provided by centralization through the tailoring of policies for individual colonies. This policy led to the adoption of the Quebec Act of 1774 which tolerated Catholicism in Quebec and the feudal institutions inherited from the French while also putting a barrier to westwards settlement. That the Quebec Act was known in the American colonies as one of the Intolerable Acts should be sufficient to understand the reaction. Through this indirect channel, the American Revolution was initiated.

1.5 Conclusion

This article has argued that a public choice interpretation can be made of the causes that led to the American Revolution. Indeed, one of the key factors in the revolution according to historians like Greene (1986, 1994, 2000) was the attempt on the part of Britain to provide a more centralized management of the Empire. This policy shift which was based on the fear that "the extensive autonomy enjoyed by the colonies might somehow lead to their loss" (Greene 2000, 99). It was heavily informed by the costly venture of expelling the Acadians in the Maritime provinces of Canada, an operation of minor strategic importance. The colonists reaped the gains of the expulsion while the fiscal responsibility was shouldered by London. The virtual autonomy over internal affairs granted by the policy of salutary neglect enabled the rent-seeking behavior of the colonists who wished to obtain the farmlands of the Acadians. The policy of centralization was a response to this issue since it limited the autonomy of the colonies and also permitted the adoption of custom-made policies centrally designed for each colony. That policy was also

the result of rent-seeking as the actors who designed them were responding to a growing constituency of government creditors, many of which were members of the government concerned with default risks.

As such, this article provides a public choice interpretation of one of the (many) leading causes of the American Revolution.

Acknowledgements The author wishes to acknowledge the help of Alex Salter, Adam Martin, Manuel Bautista, Art Carden, John Lovett and Joshua Hall. The Free Market Institute at Texas Tech provided financial support.

References

Akins T (1869) Public documents on the province of Nova Scotia. Charles Annand Publishers, Ann Arbor

Allen R (1992) His majesty's Indian allies: British Indian policy in the defence of Canada, 1774–1815. Dundurn Press, Toronto

Bordo M, Cortés-Conde R (2001) Transferring wealth and power from the Old and New World. Cambridge University Press, Cambridge

Broadberry S, Campbell BM, Klein A, Overton M, van Leeuwen B (2015) British economic growth, 1270–1870. Cambridge University Press, Cambridge

Clark G (2001) Debt, deficits, and crowding out: England, 1727–1840. Eur Rev Econ Hist 5(3):403–436

Davis L, Huttenback R (1982) The cost of Empire. In: Ransom R, Sutch R, Walton G (eds) Explorations in the new economic history. Academic, New York

de Figueiredo Jr R, Rakove J, Weingast B (2006) Rationality, inaccurate mental models, and self-confirming equilibrium: a new understanding of the American Revolution. J Theor Polit 18(4):384–415

Eccleshall R, Walker G (1998) Biographical Dictionary of British prime ministers. Routledge, London

Faragher JM (2005) A great and noble scheme: the tragic story of the expulsion of the French Acadians from their American homeland. W.W. Norton and Company, New York

Faragher JM (2006) A great and noble scheme: thoughts on the expulsion of the Acadians. Acadiensis 36(1):82–92

Geloso V (2015) Toleration of Catholics in Quebec and British public finances, 1760 to 1775. Essays Econ Bus Hist 33:51–80

Greene J (1986) Peripheries and center: constitutional development in the extended polities of the British Empire and the United States, 1607–1788. University of Georgia Press, Athens

Greene J (1994) Negotiated authorities: Essays in colonial political and constitutional history. Rutgers University Press, New Brunswick

Greene J (2000) The American Revolution. Am Hist Rev 105(1):93–102

Griffiths N (1992) The contexts of Acadian History, 1686–1784. McGill-Queen's University Press, Kingston

Gwyn J (1980) British government spending and the North American colonies 1740–1775. J Imp Commonw Hist 8(2):74–84

Gwyn J (1998) Excessive expectations: maritime commerce and the economic development of Nova Scotia, 1740–1870. McGill-Queen's University Press, Kingston

Hully T (2012) The British Empire in the Atlantic: Nova Scotia, the board of trade, and the evolution of imperial rule in the mid-eighteenth century. PhD dissertation, University of Ottawa, Ottawa

Johnston A (2003) Borderland worries: loyalty oaths in Acadie/Nova Scotia, 1654–1755. Fr Colon Hist 4:31–48
Krueger A (1974) The political economy of the rent-seeking society. Am Econ Rev 64(3):291–303
Lawson P (1989) The imperial challenge: Quebec and Britain in the age of the American Revolution. McGill-Queen's University Press, Kingston
Mitchell B (1988) British historical statistics. Cambridge University Press, Cambridge
O'Brien P (1988) The political economy of British taxation, 1660–1815. Econ Hist Rev 41(1):1–32
O'Brien P (2011) The nature and historical evolution of an exceptional fiscal state and its possible significance for the precocious commercialization and industrialization of the British economy from Cromwell to Nelson. Econ Hist Rev 64(2):408–446
Plank G (2003) Unsettled conquest: the British campaign against the peoples of acadia. University of Pennsylvania Press, Philadelphia, PA
Rabushka A (2008) Taxation in colonial America. Princeton University Press, Princeton, NJ
Shortt A, Doughty J (1918) Documents relating to the constitutional history of Canada, 1759 to 1791. Board of Historical Documents, Ottawa
Speck W (1994) The international and imperial context. In: Greene J, Pole J (eds) Colonial British America: essays in the new history of the early modern era. Johns Hopkins University Press, Baltimore
Stanbridge K (2003) Quebec and the Irish catholic relief act of 1778: an institutional approach. J Hist Soc 16(3):375–404
Stasavage D (2003) Public debt and the birth of the democratic state: France and Great Britain 1688–1789. Cambridge University Press, Cambridge
Stasavage D (2007) Partisan politics and public debt: the importance of the 'Whig Supremacy' for Britain's financial revolution. Eur Rev Econ Hist 11(1):123–153
Tullock G (1967) The welfare costs of tariffs, monopolies, and theft. Econ Inq 5(3):224–232

Chapter 2
North-South Alliances During the Drafting of the Constitution: The Costs of Compromise

Robert A. McGuire

Abstract This chapter challenges the long-standing conclusion that North-South alignments helped bring the 1787 Constitutional Convention to a successful conclusion. The widely divergent economic interests between the regions regarding commercial and merchant activities, imports and exports, and slavery and the slave trade created such widely divergent sectional differences that the North-South agreements and compromises that were necessary to complete the Constitution created a governing institution that sowed the seeds of its own downfall. By 1861, the Constitution's original design could no longer serve as the nation's governing institution; its design created circumstances that led to southern secession and a civil war that killed and wounded more than a million Americans, cost several billion dollars, and required three major amendments to "save" the Constitution as the nation's governing institution. This chapter draws on economic reasoning, political theory, and the historical record of the 1787 Constitutional Convention to challenge the long-standing conclusion that the North-South alignments helped bring the convention to a successful conclusion. The methodological approach involves juxtaposing economic principles and the issue positions of the framers and their states on the major North–South agreements and compromises among the delegates.

2.1 Introduction

> The great danger to our general government is the great southern and northern interests of the continent, being opposed to each other. – James Madison, Constitutional Convention, June 29, 1787 (Farrand 1911a, p. 476)
> We must make concessions on both sides. Without these the constitutions of the several states would never have been formed – Elbridge Gerry, Constitutional Convention, July 2, 1787 (Farrand 1911a, p. 515)

R. A. McGuire (✉)
The University of Akron, Akron, OH, USA
e-mail: rmcguire@uakron.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_2

During the summer of 1787, delegates appointed by the legislatures of twelve of the thirteen American states attended a convention of the states to consider revising the Articles of Confederation and Perpetual Union. Given that many of America's political leaders believed an economic and political crisis existed in the nation under a flawed Articles of Confederation, the delegates ultimately drafted an entirely new governing document. They drafted the U.S. Constitution.

At the time of the Constitutional Convention, there had emerged three regions of differing economic development in the nation due in large measure to market forces and differences in factor endowments. New England and the middle states each specialized in their own relatively broad mix of goods and services while the South focused on the production of a limited number of goods. In New England, extractive industries, shipping and commercial services, and small scale farming, produced by free labor, predominated. In the middle states, diversified farming and grain and food production, produced by free labor, predominated. In the South, three cash crops—tobacco, rice, and indigo—produced by slaves, predominated.[1] The three regions differed not only in their mix of output and labor systems but also in the degree to which they depended on foreign trade. The highest dependency was in the South, with its exports dominating those from the other two regions. Exports per capita in the South were nearly twice those in the middle colonies and over twofold greater than in New England. Given that the South's primary export crops were produced by slaves working on plantations and receiving a subsistence income, the returns from exports accrued primarily to the white population. Exports per white person in the South were approximately three to three and a half times that generated by the white population in the middle colonies and New England.[2]

As a result of these differences, the delegates to the Constitutional Convention represented widely divergent, and often conflicting, economic and sectional interests. This necessitated the establishment of various alliances among the framers to bring the convention to a "successful" conclusion.[3] That is, to accomplish a constitutional settlement among delegates with widely divergent interests required delegates to align with others, finding common agreement, or reaching some bargain or compromise, on their conflicting interests. Among the most important agreements, bargains, and compromises at the convention were those that involved alliances among northern (eastern) and southern delegates.

This chapter examines the major North-South alliances that are reputed to have taken place at the Constitutional Convention, challenging a long-standing consensus in the literature that the agreements and compromises between northern and southern delegates on several issues brought the convention to a "successful" conclusion.

[1] For an excellent discussion of the development of regional specialization and trade, see Shepherd and Walton (1972, Chapters 2 and 3).

[2] Exports per capita and per white resident are taken from Table 3.2 in Shepherd and Walton (1972, p. 47).

[3] For examples of this consensus view, see most any introductory American government, history, or economic history textbook. Also, see Farrand (1904) and Finkelman (1987, 2002).

While it is true that a constitutional settlement in 1787 created a new government, the contention here is that the widely divergent economic interests existing at the time created such divergent sectional conflicts that the various agreements, bargains, and compromises between northern and southern delegates, which were necessary for agreement on a constitution, created a governing institution that sowed the seeds of its own destruction. Specifically, it is argued that the *original* design of the Constitution created circumstances that led to southern secession and a civil war that killed and wounded more than a million Americans, cost several billion dollars, and required three major amendments to preserve the Constitution as the nation's governing institution.

The chapter draws on economic reasoning, political theory, and the historical record of the 1787 Constitutional Convention for its support. The methodological approach involves juxtaposing economic principles and a qualitative examination of the issue positions (not the strategies) of the framers and their states on the reputed major agreements, bargains, and compromises among northern and southern delegates. The issue positions on the following are examined:

1. The compromise among northern and southern delegates to count slaves as three-fifths of other persons for purposes of representation and direct taxation.
2. The decision to use electors to select the President, rather than appointment by the national legislature or selection by the people, which, it is claimed, is a proslavery (pro-South) provision.
3. The framers' decision to include a fugitive slave clause in the Constitution, an obvious proslavery (pro-South) provision.
4. The agreement between northern and southern (in particular, lower South) delegates to prohibit export tariffs but allow import tariffs.
5. The agreement between northern and lower South delegates to prohibit national interference in the international slave trade for the first twenty years of the Constitution's operation (rather than a shorter proscription or none).
6. The bargain between northern and lower South delegates to reject a two-thirds congressional vote (and thus adopt a simple-majority rule) to enact commercial regulations (navigation acts), including taxation of imports.

The principal conclusion drawn is that the reputed North-South alliances at the Constitutional Convention on many of the major issues imposed such high costs (most of which were implicit) on the nation during the ensuing decades (costs which were undoubtedly unforeseen and unintended) that by 1861 the Constitution's original design could no longer serve as the nation's governing institution. So problematic was the U.S. Constitution that the South eventually designed its own constitution to serve as the governing document for a new confederation of states in 1861, the Confederate State of America.

2.2 The Political Economy of Constitution Making

A key question in constitutional political economy—the study of the choice of the rules governing collective behavior—is how do rational actors make choices regarding the basic design of a constitution—that set of fundamental rules which is agreed on in advance, and within which future collective decisions will be made. The primary conclusion in this literature is that a self-interested actor behaves differently during the constitutional stage of collective decision making than during any other stage. The presumption is that partisan interests do not influence the basic design of a constitution.

In their pioneering formulation of a theory of constitutional choice, Buchanan and Tullock (1962, Chapter 6) describe two levels of collective decision making—the "constitutional level" and the "operational level." The constitutional level is where individuals choose a set of general rules to be used for all future collective choices. The operational level corresponds to choices of laws, regulations, and statutes, *not* general decision-making rules. According to Buchanan and Tullock (1962, p. 78), at the constitutional level of collective choice an individual "will not find it advantageous to vote for rules that may promote sectional, class, or group interests because by presupposition, he is unable to predict the role that he will be playing in the actual collective decision-making process . . . in the future." Constitutional actors behave as if they are behind a "veil of uncertainty." The discussion among individuals about constitutional design will proceed without intense conflicts about the general effects or workings of rules, because it "should not be unlike that of possible participants in a game . . . Since no player can anticipate which specific rules might benefit him during a particular play of the game . . . [they] attempt to devise a set of rules . . . for the average or representative player" (Buchanan and Tullock 1962, pp. 79–80). Constitutional participants likewise will "take a position as 'a representative' or 'randomly distributed' participant", choosing "the best set of rules for the social group"—an efficient constitution (Buchanan and Tullock 1962, p. 96). Moreover, because constitutional actors are behind a veil of uncertainty, "the purely selfish individual and the purely altruistic individual may be indistinguishable in their behavior" (Buchanan and Tullock 1962, p. 96). Behind the veil of uncertainty, both would design a constitution that is in society's interest because disinterested behavior would predominate. It follows that partisan interests do *not* matter for constitutional choice.

At the operational level of collective decision making, individuals make collective choices within certain agreed-on rules. At this level, the individual calculus is similar to deciding on the allocation of resources within a well-defined time period where the individual knows his interests and position relative to others. As noted, this level corresponds to the choice of laws, regulations, and statutes. As Buchanan and Tullock (1962, p. 120) maintain, at the operational level, the rational individual "is assumed to be motivated by a desire to further his own interest, to maximize his expected utility, narrowly or broadly defined . . . the individual's interest will be more readily identifiable and more sharply distinguishable from those of his fellows

than was the case at the constitutional level of decision." It follows that partisan interests *matter* for operational issues, which is consistent with the behavior of political actors in modern political agency theory, where politicians are the agents of the principals (constituents) who elect and/or support them and are presumed to be motivated by partisan behavior.

Buchanan and Tullock (1962, p. 80), however, maintain that their theory of constitutional choice will *not* hold if there exists "a sharp cleavage of the population into distinguishable social classes or ... groupings," individual members of a dominant interest group "would never rationally choose to adopt constitutional rules giving less fortunately situated individuals" equal standing. Buchanan and Tullock (1962, pp. 78–79) also argue that if the veil of uncertainty is not sufficiently thick a "rational utility-maximizing individual will support the adoption of rules designed specifically to further partisan interests" and oppose the adoption of efficient constitutional rules. They suggest that an individual can see through the veil if (1) he is able to predict which issues will arise under alternative rules, (2) he can determine the outcomes under the alternative rules, (3) he prefers an outcome under an inefficient rule over all other outcomes under the efficient rule, and (4) all individuals agree to the inefficient rule (Buchanan and Tullock 1962, p. 79).

In subsequent work, Buchanan (1975) goes beyond the highly stylized model of Buchanan and Tullock (1962) in an attempt to develop a more general model of constitutional contract. He argues that "allowance should be made for the existence of substantial differences among persons in the original ... setting" (Buchanan 1975, p. 54), and then models how rational, self-interested actors are willing to form a constitution out of anarchy. Behind a veil of uncertainty individuals are able to see that rules which support the gains from direct production, specialization, and exchange are in their interest. Yet, with substantial inequality, anarchy *could* be superior to direct production and exchange. In this case, constitutional agreement is still possible but existing wealth differences must first be ameliorated through redistribution of wealth among the society's members (Buchanan 1975, Chapter 4). Brennan and Buchanan (1985) offer a positive constitutional theory that provides a self-interest explanation for choosing nonpartisan "rules-of-the-game"—that is, for choosing a "just" constitution. But substantial differences among constitutional actors are also problematic in their model, creating a probable roadblock to nonpartisan constitutional settlement.

Buchanan and Vanberg (1989) and Buchanan (1991a) argue that there is still a lack of conflict about the expected workings of different rules among constitutional actors, because in the traditional contractarian approach to constitutional choice conflict-free discussion is either an underlying assumption of the model (Buchanan and Tullock 1962), or conflict is eliminated entirely by assuming that individuals are fully informed about the general effects of constitutional rules (Rawls 1971). Accordingly, when modeling behavior at the 1787 Constitutional Convention, Buchanan and Vanberg (1989, pp. 17–18; 23–25) presume a veil of uncertainty to eliminate conflicts over partisan interests (and thus *presume* disinterested behavior), which allows them to model an equilibrium outcome in which a constitution

contains both fair and efficient rules—"fair" in terms of the distribution of the effects of rules and "efficient" in terms of the effects of rules on the level of welfare.

Yet Buchanan (1991b, p. 56) recognizes the view in the political rent-seeking literature that "identifiable and conflicting constitutional interests will prevent . . . [rational actors] from actually realizing the potential gains from constitutional cooperation" when placed in real world settings. While Buchanan recognizes that widely divergent interests can make the veil of uncertainty transparent in real world contexts, he stresses that the characteristics of any actual constitutional setting, its institutional structure, determines the actual thickness of the veil. Moreover, he suggests that rational actors might recognize this "rent-seeking trap" and deliberately structure a constitutional setting so as to thicken the veil, but Buchanan offers no evidence of a real world setting in which this actually takes place (Buchanan 1991b, p. 56).

Many constitutional political economists unquestioningly accept the presumption that the nation's Founding Fathers were disinterested when they drafted the U.S. Constitution. Buchanan and Vanberg (1989) explicitly model the behavior of James Madison and those that followed his leadership at the Constitutional Convention as a real-world example of constitutional economics, including the presumption of disinterestedness. In another study of James Madison, Dorn (1991) presumes that disinterested behavior predominated in 1787. Lowenberg and Yu (1992, p. 61) argue that because a viable "exit" option existed for many of the parties at the Constitutional Convention—they claim that secession was feasible for several states—the convention delegates would have exhibited disinterested behavior. These studies, though, do not provide any historical support for a presumption of disinterested behavior at the Constitutional Convention; they simply assume the convention was a "constitutional" setting and thus disinterested behavior prevailed.

In contrast, economic historians and political scientists have empirically addressed the issue of interested versus disinterested behavior at the Constitutional Convention. In a series of studies in the 1980s and 1990s, McGuire (1988) and McGuire and Ohsfeldt (1984, 1986, 1997), culminating in McGuire (2003), conclude that the convention does not fit neatly into either the constitutional or the operational level of collective decision making. The rules and structure of the convention and the concentrated wealth holdings of many of the delegates suggest that the actual historical setting was different from a pure constitutional or operational setting. According to McGuire and McGuire and Ohsfeldt, the convention was a combination of both levels of collective choice. In addition to considering general decision-making (constitutional) rules, the convention considered numerous interest-specific (operational) rules. As a result, McGuire (2003) concludes that interested partisan behavior existed on various issues at the convention. Jillson (1981, 1988) and Jillson and Eubanks (1984) also argue for the importance of interested behavior, at least on some issues and during particular times at the convention. Jillson (1981) formally examines the voting alignments among the states, indicating that most of the alignments long recognized by historians existed on those issues for which they were expected. Expanding on this approach, Jillson and Eubanks (1984) identify what they refer to as

"higher order" issues and "lower order" issues at the convention. They formally identify voting alignments among state delegations, concluding that the convention progressed from higher order to lower order issues and the votes on issues of basic constitutional design (higher order issues) were decided on the basis of principle; whereas votes on more interest specific issues (lower order issues) were decided by specific economic and political interests. Accordingly, the convention progressed from less-interested behavior in its earlier deliberations to more-interested behavior in its later deliberations (see, also, Jillson 1988).

The question for the present study is to what extent disinterested or "constitutional" behavior actually prevailed during the drafting of the U.S. Constitution. First, the available evidence indicates that at the time of the Constitutional Convention:

1. there was "a sharp cleavage of the population into distinguishable social classes or ... groupings;"
2. there were "substantial differences among persons;" and
3. the framers of the Constitution represented "identifiable and conflicting constitutional interests."

In short, widely divergent economic and sectional interests were represented at the convention. As a result, any veil of uncertainty that the framers might have been behind would have been fairly transparent. Consequently, modern constitutional political economy argues that rational utility-maximizing actors (both northern and southern delegates) would have supported "the adoption of rules designed specifically to further partisan interests." Second, the existing evidence indicates that interested, partisan behavior on the part of the framers in fact took place during much of the convention; the convention delegates in fact attempted to adopt constitutional rules that furthered their own economic and sectional interests (Jillson 1988; McGuire 2003; Robertson 2005). The contention here, though, is that the southern framers (in particular, those from the lower South) failed in their attempt to further their section's (long run) interests during the drafting of the Constitution.

2.3 The North-South Alliances at the 1787 Constitutional Convention

That the South was, compared to the North, overwhelmingly agrarian and dependent on plantation slavery, and more export-based, was at the heart of the sectional differences in late eighteenth-century America. And these differences necessitated several alliances of northern and southern delegates on various issues at the Constitutional Convention so a constitution could be drafted and approved by the states. At the convention, nearly all North-South disputes involved some aspect of slavery and/or navigation acts (commercial regulations). Among the disputed issues was how slaves would be counted in apportionment and whether slaves would be subject to taxation and, if so, how would they be taxed. Another issue was whether

slave importation would be allowed and, if so, whether such imports would be taxed. Other North-South disputes at the convention concerned whether trade (exports and imports) in general would be taxed and whether a simple or super congressional majority would be required to enact navigation acts (commercial regulations). Some scholars have argued there were sectional differences on how the executive should be chosen; also there might have been different views on how runaway slaves should be handled. In short, had some northern and southern delegates not aligned on such issues, the Constitution likely would not have been drafted nor approved by the states.

2.3.1 Apportionment and Direct Taxation

Among the most contentious and longest debated issues at the convention was the issue of the apportionment of representation among the several states and the link between taxation and representation (Article I, Section 2, Clause 3). The most obvious, though unattainable, apportionment rule that would have benefited northern (eastern) states would be to not count slaves at all; the most beneficial apportionment rule from the South's perspective, of course, would be to count slaves as equal to the white population. Yet confederation among the northern and southern states would have never taken place with either of these extreme apportionment rules; the delegates certainly knew that going in. Moreover, under either extreme apportionment rule, if taxation was to be proportional to representation, the most beneficial apportionment rule to either section of the country would always be the most costly in terms of taxation. The delegates faced the twin dilemmas of representation and taxation in the summer of 1787.

The first mention of the apportionment issue was on May 29 when Edmund Randolph (Virginia) delivered to the convention a plan for a new national government (the Virginia Plan), which included a resolution "that the rights of suffrage in the National Legislature ought to be proportioned to the Quotas of contribution, or to the number of free inhabitants, as the one or the other rule may seem best in different cases" (Farrand 1911a, pp. 18–23).[4] When the delegates took up the apportionment resolution the next day, with the convention organized as a Committee of the Whole House, James Madison (Virginia) quickly made it clear there might be North-South problems on apportionment, as he warned that the clause "or to the number of free

[4]Unless otherwise noted, all quotations of speeches at the Constitutional Convention are from James Madison's "Notes" contained in Farrand (1911a,b) and are Madison's synopses of what was said at the convention. It needs to be noted that Madison made various changes to his notes after they were initially transcribed; some of these were made as much as forty or more years later. All such changes "are indicated by enclosing them within angle brackets ⟨⟩" (Farrand 1911a, p. xix). All abbreviations, punctuation, and spelling in the quotes are exactly as found in Farrand. In cases where a quotation is from another delegate's notes, that delegate's name is also listed in the parentheses.

inhabitants" should be deleted from the resolution because it might lead to debate that would divert the delegates from the main question of whether suffrage should be changed from the existing Confederation system (Farrand 1911a, pp. 35–38). But Madison's fears were avoided, as the issue was postponed.

Debate on apportionment began in earnest on June 9 (Farrand 1911a, pp. 176–180) and continued on June 11 (Farrand 1911a, pp. 196–201), when the well-documented small versus large state differences, which were often intertwined with North-South differences, began. Among the more significant motions, Rufus King (Massachusetts) and James Wilson (Pennsylvania) on June 11 moved "that the right of suffrage in ⟨the first branch of⟩ the national Legislature ought not to be according the rule established in the articles of Confederation, but according to some equitable ratio of representation" (Farrand 1911a, p. 196), which passed in the affirmative. After which, Wilson moved, and Charles Pinckney (South Carolina) seconded, that the following be added to the end of the previous motion: "in proportion to the whole number of white & other free Citizens & inhabitants of every age sex & condition including those bound to servitude for a term of years and three fifths of all other persons not comprehended in the foregoing description, except Indians not paying taxes, in each State" (Farrand 1911a, p. 201). Wilson noted at the time of his motion that "this mode had already received the approbation of eleven states in their acquiescence to the quota made by congress" (Farrand 1911a, p. 205; Robert Yates' notes). Thus, the first mention of the three-fifths rule took place early in the convention.

Wilson and Pinckney's motion passed in the affirmative, 9-2, in the Committee of the Whole on June 11, with only New Jersey and Delaware in opposition.[5] As controversial as the three-fifths rule would be later in the convention, only Elbridge Gerry (Massachusetts) spoke on the motion, arguing that "The idea of property ought not to be the rule of representation. Blacks are property, and are used to the southward as horses and cattle to the northward; and why should their representation be increased to the southward on account of the number of slaves, than horses or oxen to the north?" (Farrand 1911a, pp. 205–206; Robert Yates' notes). No other delegate even mentioned sectional issues until June 26, when Charles Pinckney's older cousin, Charles Cotesworth (C.C.) Pinckney (South Carolina), in discussing another resolution, noted that the states "had different interests. Those of the Southern, and of S. Carolina in particular were different from the Northern" (Farrand 1911a, p. 421).

Debate on apportionment did not resume again until June 27. And after 4 days of debates on the benefits of equal versus proportional suffrage schemes to small versus large states, James Madison on June 30:

[5]The vote, though, was in no sense a decisive vote on the three-fifths rule as it was a vote of the Committee of the Whole to consider the resolution put forth by Randolph; it was not a binding vote or the final vote of the convention. As a Committee vote, it simply meant that the approved resolution would afterward be presented to the convention to consider. And the rules of the convention allowed individual framers freedom to propose and discuss any issue at the convention, including the ability to reconsider previously debated issues.

> contended that the States were divided into different interests not by their difference of size, but by other circumstances; the most material of which resulted partly from climate, *but principally from . . . their having or not having slaves*. These two causes concurred in forming the great division of interests in the U. States. It did not lie between the large & small States: it lay between the Northern & Southern . . . instead of proportioning the votes of the States in both branches, to their respective numbers of inhabitants computing the slaves in the ratio of 5 to 3. they should be represented in one branch according to the number of free inhabitants only; and in the other according to the whole no. counting the slaves as ⟨if⟩ free." (Farrand 1911a, pp. 486–487; emphasis added).

In the next meeting of the convention on July 2, Charles Pinckney agreed that there were major differences between northern and southern states, arguing that "an equality of votes in the 2d. branch inadmissible" (Farrand 1911a, pp. 510–511). Given the continuing differences on apportionment, the convention appointed a grand committee of all the states to consider the proper representation in the two branches of the general government.

On July 5, the grand committee delivered its report which proposed that each state's representation in the first branch be based on population, that money bills originate in the first branch, and that representation in the second branch be equal. In response, the delegates began what became more than a week of debates on the apportionment issue and representation of slaves. Several delegates began arguing that property or wealth should be a basis for representation, particularly in the second branch. Pierce Butler (South Carolina) was the first to urge "that the 2d. branch ought to represent the States according to their property" (Farrand 1911a, p. 529). Although Gouverneur Morris of Pennsylvania "thought property ought to be taken into the estimate as well as the number of inhabitants" (Farrand 1911a, p. 533), most proponents of apportionment based on property, wealth, or, as John Rutledge (South Carolina) proposed, "the sums paid towards the general revenue by the inhabitants of each state" (Farrand 1911a, p. 534) were southerners, particularly from the lower South, especially South Carolina. And while Charles Pinckney argued the next day that "The number of inhabitants . . . [were] the only just & practical rule" for the first branch, he added that "the blacks ought to stand on an equality with whites" (Farrand 1911a, p. 542).

On July 9, a subcommittee proposed to initially fix the number of representatives for each state, which would total to 56, for the first meeting of the first branch, but allow future adjustments based on population and wealth. In reaction, delegates questioned how the numbers were determined and what weights were attached to the number of inhabitants versus wealth. Among them, William Paterson (New Jersey) was against a "Combined rule of numbers and wealth, as too vague. For this reason N. Jersey was agst. it" (Farrand 1911a, p. 561). Moreover, he "could regard negroes slaves in no light but as property" and given that "Negroes are not represented in the States to which they belong, why should they be represented in the Genl. Govt." (Farrand 1911a, p. 561).

At the next meeting on July 10, a report of the committee of all the states proposed to raise the number of representatives in the House from the proposed 56 to 65, with five more going to northern states, three more going to the upper

South, and only one more representative going to lower South states (Georgia). The lower South delegates complained about insufficient representation; the South Carolina delegates, John Rutledge and C.C. Pinckney, in fact, moved that New Hampshire's number of representatives be reduced from 3 to 2, but the motion failed in the Committee of the Whole House. C.C. Pinckney and Alexander Martin (North Carolina) then moved that North Carolina's representation be increased from 5 to 6, which also failed. C.C. Pinckney and Pierce Butler next moved that South Carolina's representation be increased from 5 to 6, but it also failed. After which, C.C. Pinckney and William Houstoun (Georgia) moved that Georgia's representation be increased from 3 to 4, which likewise failed. The losing side in each case never received more than the votes of four states and as few as two. And only two lower South states—Georgia and South Carolina—voted yes on all four motions and against the grand committee's overall motion (Farrand 1911a, pp. 568–570).

The convention resumed debate the next day on an earlier motion made by Edmund Randolph requiring the legislature to take a periodic census so that future representation can be adjusted. George Mason (Virginia) made clear that he "considered a Revision from time to time according to some permanent & precise standard as essential to ye. fair representation required in the 1st. branch" (Farrand 1911a, p. 578). In Mason's view, southern states needed security for their future representation because those with power will never give it up, so that in the future "If the S. States therefore should have 3/4 of people of America within their limits, the Northern will hold fast the majority of Representatives. 1/4 will govern the 3/4" (Farrand 1911a, p. 578). Mason concluded that blacks must be represented. Hugh Williamson (North Carolina) then moved to postpone Randolph's motion to consider an alternative that, "in order to ascertain the alterations that may happen in the population & wealth of the several states," required an annual census "of the free white inhabitants and frac35 of those of other descriptions" (Farrand 1911a, p. 579). In reaction, C.C. Pinckney and Butler "insisted that blacks be included in the rule of Representation, equally with the Whites" and "moved that the words 'three fifths' be struck out" (Farrand 1911a, p. 580).[6] Elbridge Gerry responded that three-fifths "was to say the least the full proportion that could be admitted" (Farrand 1911a, p. 580). His fellow Massachusetts delegate, Nathaniel Gorham, reminded the convention that Congress had previously fixed the three-fifths ratio "as a rule of taxation" (Farrand 1911a, p. 580). Butler "insisted that the labour of a slave in S. Carloina was as productive & valuable as that of a freeman in Massts. ... consequently an equal representation ought to be allowed" (Farrand 1911a, p. 580). George Mason (an upper South delegate), while arguing that slaves ought to be counted in representation, "could not however regard them as equal to freemen and could not vote for them as such" (Farrand 1911a, p. 581). When the question on the motion to strike out "three fifths" was called in the Committee of the Whole House, it failed, 3-7, as only Georgia, South Carolina, and Delaware voted for it (Farrand 1911a, p. 581).

[6]This begins intense debate on the three-fifths rule, which was agreed to 2 days later on July 13.

Gouverneur Morris expressed objections to Williamson's motion arguing that, among others, "It fettered the legislature too much" and, if the census was to ascertain changes in "population & wealth," then the motion was inconsistent because "If slaves were to be considered as inhabitants, not as wealth, then the sd. Resolution would not be pursued: If as wealth, then why is no other wealth but slaves included?" While amendments may remove these objections according to Morris, his primary "objection was that the number of inhabitants was not a proper standard of wealth" (Farrand 1911a, pp. 581–582). John Rutledge then moved that "the Legislature shall proportion the Representation according to the principles of wealth & population" every year (Farrand 1911a, p. 582). Debate ensued on whether wealth should be counted, how it should be measured, and whether the resolution was too impractical or shackled the legislature too much. Gouverneur Morris did not believe "that numbers would be a just rule at any time" and another objection he had with "admitting the blacks into the census, was that the people of Pena. would revolt at the idea of being put on a footing with slaves. They would reject any plan that was to have such an effect" (Farrand 1911a, p. 583). In a lengthy response, James Madison "could not agree that any substantial objection lay agst. fixing numbers for the perpetual standard of Representation. It was said that Representation & taxation were to go together; that taxation & wealth ought to go together, that population and wealth were not measures of each other" (Farrand 1911a, p. 585). Madison admitted that numbers of inhabitants were not an accurate measure of wealth but "in the U. States it was sufficiently so for the object in contemplation" (Farrand 1911a, p. 585). George Mason repeated his opinion that the proposed constitution must include a mechanism to adjust future representation to protect southern and western interests from minority domination in the future (Farrand 1911a, p. 586). The Committee then agreed to the first clause of Williamson's motion, "as to taking a census of free inhabitants" and took up "the next clause as to & frac35; of the negroes (Farrand 1911a, p. 586).

Rufus King said that he was "much opposed to fixing numbers as the rule of representation, was particularly so on account of the blacks," arguing that "the admission of them along with whites at all, would excite great discontents among the States having no slaves" (Farrand 1911a, p. 586). Besides, said King, "the Southern States had received more than the number of their white & three fifths of their black inhabitants entitled them to" (Farrand 1911a, p. 586). While Roger Sherman (Connecticut) was satisfied with the proposed allotments, including representation for slaves; James Wilson "did not well see on what principle the admission of blacks in the proportion of three fifths could be explained. Are they admitted as citizens? . . . Are they admitted as property?" (Farrand 1911a, p. 587). Gouverneur Morris felt forced to do injustice to the southern states because "he could never agree to give such encouragement to the slave trade as would be given by allowing them a representation for their negroes" (Farrand 1911a, p. 588). While the second clause of Hugh Williamson's motion, to include three-fifths of blacks in the census, passed in the affirmative, 6-4 (South Carolina voted no), in the Committee of the Whole

House on July 11, the whole motion to require an annual census to ascertain any changes in the population and wealth of the states failed, 0-9 (Farrand 1911a, p. 588).

At the opening of the debates the next day, July 12, Gouverneur Morris moved to insert at the end of the apportionment clauses a "proviso that taxation shall be in proportion to Representation" (Farrand 1911a, pp. 591–592). Pierce Butler, George Mason, C.C. Pinckney, and James Wilson all saw the justice of Morris' proviso but agreed that it should be limited to "direct" taxation, which it was, as the convention unanimously agreed to insert the word "direct." Butler again said that "Representation sd. be according to the full number of inhabts. including all the blacks" (Farrand 1911a, p. 592).[7] William Davie (North Carolina) said "it was high time to speak out" because some delegates want "to deprive the Southern States of any share of Representation for their blacks," concluding that "N. Carola. would never confederate on any terms that did not rate them at least as & frac35." If not, Davie said, "the business was at an end" (Farrand 1911a, p. 592). Interestingly, similar to the earlier sentiments of Roger Sherman, another northerner, William Samuel Johnson (Connecticut), believed that "wealth and population are the true, equitable rule of representation" and with "population being the best measure of wealth" it ought to be the rule "that all descriptions including blacks equally with the whites, ought to fall within the computation" (Farrand 1911a, p. 593). Gouverneur Morris again warned that "the people of Pena. will never agree to a representation of Negroes" (Farrand 1911a, p. 583).

The debates continued throughout the day on: (1) the link between direct taxation and representation, (2) whether representation would include slaves and to what degree, and (3) whether there would be a periodic census to adjust future representation. But the battle lines had been drawn: upper South delegates insisting on slave representation and a periodic census, as Edmund Randolph said, "rating the blacks at & frac35; of their number" because "security ought to be provided for including slaves in the ratio of representation ...the holders of it [such a species of property] would require this security" (Farrand 1911a, p. 594); some northerners (easterners), particularly Roger Sherman and his fellow Connecticut delegates, supporting slave representation, supported counting them equally; the lower South delegates insisting on, as Charles Pinckney moved one more time, counting "blacks equal to the whites in the ratio of representation" (Farrand 1911a, p. 596); and other northerners (easterners), in particular Massachusetts and Pennsylvania delegates—the most vocal being Gouverneur Morris—opposing any representation for the slaves and demanding direct taxation in proportion to representation.[8] By the end of the day, the whole proposition, "as proportioning

[7]Interestingly, for the first time in the convention, the southern view that exports should not be taxed was expressed explicitly: C.C. Pinckney said that he hoped a clause would be inserted in the proposed constitution that prohibited export taxes (Farrand 1911a, p. 592).

[8]Pinckney's July 12 motion to count "blacks equal to the whites" was defeated, again with only the two lower South states, Georgia and South Carolina, voting for the motion (Farrand 1911a, p. 596).

representation to direct taxation & both to the white & frac35; of black inhabitants, & requiring a census within six years–& within every ten years afterwards," passed in the affirmative, 6-2-2 (with South Carolina divided), in the Committee of the Whole House (Farrand 1911a, p. 597).

The next day, July 13, the convention took up the issue of raising monies for the public treasury by direct taxation prior to a census and linking the amount to the number of representatives. Again, much of the debate focused on the number of representatives initially apportioned to each state and its relationship to the slave population of the states. After Elbridge Gerry's motion that prior to the first census "all monies for supplying the public treasury by direct taxation shall be raised from the several states according to the number of their representatives respectively in the 1st. branch" (Farrand 1911a, p. 603) was approved, the Committee of the Whole House reconsidered the apportionment resolution that was approved the day before, on a motion of Edmund Randolph, "in order to strike out 'Wealth' and adjust the resolution to that requiring periodical revisions according to the number of whites & three fifths of the blacks" (Farrand 1911a, p. 603). Which led Gouverneur Morris again to oppose the proposed apportionment rule, as he complained that "If Negroes were to be viewed as inhabitants . . . they ought to be added in their entire number . . . If as property, the word wealth was right, and striking it out would produce the very inconsistency which it was meant to get rid of." Now, the direction of the debates led him "into deep meditation" from which he concluded "A distinction had been set up & urged, between Nn. & Southn. States." He still considered the differences to be groundless, but saw that it persisted and "the Southn. Gentlemen will not be satisfied unless they see the way open to their gaining a majority," which he feared could lead to entanglements with Spain over the Mississippi. So he wanted "to know what security the Northn. & middle States will have agst. this danger" (Farrand 1911a, pp. 603–604).

Despite Morris' continuing and vociferous objections, the motion "to strike out wealth & make the change as moved by Mr. Randolph" passed in the affirmative, 9-0-1, on July 13 (Farrand 1911a, p. 606), which essentially settled the issue of apportionment in the House and the link between direct taxation and representation—three-fifths of slaves would be counted for purposes of representation and direct taxation. But not all parts of the Grand Compromise had been settled. The compromise on origination of money bills in the House and equal representation in the Senate remained. While the proceedings continued to be quite contentious, both of these issues were settled within the next two meetings when on July 16 the entire Grand Compromise was passed in the affirmative.

What is clear from this examination of the positions of the delegates on the apportionment issue is that most of the delegates from the lower South states of Georgia and South Carolina did not consider the "grand compromise" to be in the interest of their states nor did many of the Massachusetts and Pennsylvania delegates consider it in their states' interest. In fact, of the South Carolina delegates that spoke at the convention, all of them were adamantly opposed to the "grand

compromise." And of the Massachusetts and Pennsylvania delegates that spoke, they were as adamantly opposed to giving representation for slaves. Consequently, given the lower South's adamant opposition, it is fair to conclude that the "grand compromise" can be considered to have been "costly" to the lower South states.

2.3.2 Election of Executive

It has been argued that the procedure for election of the executive with electors (Article II, Section 1, Clause 2)—what is known today as the Electoral College—was created as a proslavery, pro-South constitutional provision (Finkelman 2002). If true, this means there would have been likely North-South differences on the issue that would have required some alliance among the delegates to accomplish agreement on the procedure. The first mention of the election of the executive is May 29 when Randolph's Virginia Plan was introduced and in which resolution seven stated that "a National Executive be instituted; to be chosen by the National Legislature for the term of [____] years" (Farrand 1911a, p. 21). The subject was first discussed on June 1 (Farrand 1911a, pp. 64–69) and June 2 (Farrand 1911a, pp. 79–81), when presidential "electors" were proposed by James Wilson (Pennsylvania) and supported by Elbridge Gerry (Massachusetts), as an alternative to election by the National legislature. The appointment of electors was opposed by Hugh Williamson (North Carolina). In the Committee of the Whole, Wilson's motion lost 2-8; the Virginia Plan resolution to elect the president by the National legislature then passed 8-2.

Election of the executive was reconsidered on June 9 (Farrand 1911a, pp. 175–176) when it was moved by Gerry "that the National Executive should be elected by the Executives of the States whose proportion of votes should be the same with that allowed to the States in the election of the Senate," which was promptly and unanimously defeated (Farrand 1911a, p. 175). Then as part of the intense debate that followed the introduction of the New Jersey Plan to the convention by William Patterson (New Jersey) on June 15, Alexander Hamilton (New York) delivered a several hour speech on June 18 in which he discussed the defects of both the New Jersey and Virginia Plan and offered his own outline for a new central government, which included a supreme executive "the election to be made by Electors chosen by the people in the Election Districts aforesaid" (Farrand 1911a, p. 292); the second time that presidential electors were suggested in the convention and each time by northern delegates.

The election of the executive was not taken up again for a month. Then when it was considered on July 17 Gouverneur Morris stated that he "was pointedly agst. his being so chosen. He will be the mere creature of the Legisl. ... He ought to be elected by the people at large, by the freeholders of the Country" (Farrand 1911b, p. 29). Roger Sherman and Charles Pinckney were both concerned that if selected by the people the most populous states would combine to elect who they wished

and, thus, supported selection by the legislature. The debate on the election of the executive continued for the next week and a half until July 26 when the convention's proceedings through that day were committed to the Committee of Detail. During the debates on the executive on July 19, Gouverneur Morris and James Wilson continued to support selection by the people and oppose selection by the (national) legislature, Rufus King and William Patterson thought electors might be a solution, Edmund Randolph favored selection by the legislature, and James Madison opposed selection by the legislature as he also believed that selection by the people was ideal, but "There was one difficulty however of a serious nature attending a immediate choice by the people. The right of suffrage was much more diffusive in the Northern than the Southern States; and the latter could have no influence in the election on the score of the Negroes. The substitution of electors obviated this difficulty and seemed on the whole to be liable to the fewest objections" (Farrand 1911b, pp. 56–57).

So, for the first time, one of the delegates suggests that electors are a pro-South (proslavery) alternative to selection by the legislature or by the people. And it is suggested by a southerner, albeit, from the upper South state of Virginia. What follows over the next few days, though, tells the complete story of the support for (and opposition to) what became the Electoral College. Oliver Ellsworth with the support of Elbridge Gerry moved on July 19 "to strike out the appointmt. by the Natl. Legislature, and insert 'to be chosen by electors appointed by the Legislatures of the States"' (Farrand 1911b, pp. 57–58). The only recorded comment on the motion was by John Rutledge (South Carolina) who made it clear that he "was opposed to all modes except the appointmt. By the Natl. Legislature" (Farrand 1911b, p. 57). Ellsworth's motion was divided in two with the first vote on selection by electors, the second on who selects the electors. The first part "shall ye. Natl. Executive be appointed by Electors" passed in the affirmative, 6-3-1, with the three lower South states (Georgia, North Carolina, and South Carolina) being the only no votes (Farrand 1911b, p. 58). Following passage of the second part ("the electors to be chosen by the State Legislatures"), Luther Martin (Maryland) moved that the executive be ineligible for a second term, which was seconded by Hugh Williamson (North Carolina) who commented that he possessed "no great confidence in the Electors to be chosen for the special purpose. They would not be the most respectable citizens . . . They would be liable to undue influence" (Farrand 1911b, p. 58).

On July 20, George Mason supported impeachment of the executive because "One objection agst. Electors was the danger of their being corrupted by the Candidates: & this furnished a peculiar reason in favor of impeachments whilst in office" (Farrand 1911b, p. 65). Mason said he favored appointment by the (national) legislature. On July 23, a motion by Georgia's William Houstoun and North Carolina's Richard Dobbs Spaight to reconsider the use of presidential electors because of concern about their expense passed, 7-3; interestingly, all three lower South states *favored* reconsideration and both upper South states (Maryland and Virginia) *opposed* the motion (Farrand 1911b, p. 95). On July 24, Houstoun moved, and Spaight seconded, that the executive "be appointed by the 'Natl. Legislature. <instead of 'Electors appointed by the State Legislatures" according to the last

decision of the mode>" (Farrand 1911b, p. 99). Williamson generally supported the motion. Ellsworth, Gerry, Morris, and Wilson all opposed the motion, each wanting some other method (selection by a combination of the (national) legislature and electors, election by the governors, election by lot by a small number of the (national) legislature, and selection by electors chosen by the (national) legislature, respectively) (see Farrand 1911b, pp. 101–102). But the motion passes, 7-4, again, interestingly all three lower South states *favored* the change back to selection by the national legislature and both upper South states (Maryland and Virginia) *opposed* it (Farrand 1911b, p. 101).

On July 25, Ellsworth and Gerry each moved that the executive be chosen according to their preferred method (selection by a combination of the (national) legislature and electors, and election by the governors, respectively) (Farrand 1911b, pp. 108–109). Madison again said that, even with its imperfections, he considered election by the people "the best;" though, he recognized that there was a "disproportion of <qualified voters> in the N. & S. States" and, while he thought this disadvantage would decline with time, even if the South was at a disadvantage as a southerner "he was willing to make the sacrifice" (Farrand 1911b, pp. 109–111). The executive selection debates continued through the next day, July 26, as various delegates continued in their attempts to alter the selection process or the conditions of the office to their own liking. Nonetheless, the whole resolution on the executive, which said the executive would be a single person to be chosen by the national legislature for a single seven-year term to receive a fixed compensation out of the national treasury and have power to execute the national laws and be subject to removal upon impeachment and conviction, was approved, 6-3-1. Again all states of the lower South voted for the resolution while the upper South either opposed the resolution (Maryland) or was divided on it (Virginia, with John Blair and George Mason voting yes and James Madison and George Washington voting no) (Farrand 1911b, p. 121). Just before the vote, Mason had said that "an election by the Natl. Legislature as originally proposed, was the best" (Farrand 1911b, p. 119). At the end of the day, the convention's proceedings up to that date, including the executive resolution, was committed to the Committee of Detail.

What stands out so far about the convention debates on the method for choosing the executive is that other than James Madison, only one other southerner ever spoke in support of selection by electors. (On July 25, South Carolina's Pierce Butler said that he liked best the use of electors, as long as the electors were to be chosen by the state legislatures and there was equality among the states.) No delegate at the convention other than Madison ever mentioned the issue of a disproportion of votes between northern and southern states or the issue of slaves or slavery in connection with selection of the executive. In all cases, northern (eastern) delegates proposed motions that the executive should be chosen by a method that involved electors. In all cases, votes on selection by electors were opposed by the three lower South states and supported by the two upper South states. In all cases, votes on appointment of the executive by the (national) legislature were supported by the three lower South states and opposed by the two upper South states (with the exception of one divided Virginia vote).

After the report of the Committee of Detail on August 6, the convention did not take up the method of choosing the executive until August 24 when Daniel Carroll (Maryland) "moved to strike out, 'by the legislature' and insert 'by the people'" for the selection of the executive (Farrand 1911b, p. 402). The motion, which was seconded by James Wilson, was defeated, 2-9, with no discussion whatsoever. The issue did not come up again until September 4 following a Committee of Eleven report (one of several such committees appointed to find solutions to various unresolved issues) that proposed that "Each state shall appoint in such manner as its Legislature may direct, a number of electors equal to the whole number of Senators and members of the House of Representatives' instead of appointment by the (national) legislature (Farrand 1911b, p. 497). Edmund Randolph and Charles Pinckney questioned why the mode of electing the executive was changed and wanted an explanation. George Mason objected to the change because he believed that 95% of the time the selection would end up in the Senate because no candidate would receive a majority of the electoral votes (Farrand 1911b, p. 500). Pierce Butler still had objections, but thought the new mode was better than the national legislature appointing electors. Charles Pinckney objected to the new mode, believing that it would throw "the whole appointment in fact into the hands of the Senate" and "The Electors will be strangers to the several candidates and of course unable to decide o their comparative merits" (Farrand 1911b, p. 501). Hugh Williamson questioned the new mode, as did Edmund Randolph who said that he "preferred the former mode of constituting the Executive" (Farrand 1911b, p. 502). Throughout the day, Gouverneur Morris, a member of the Committee, repeatedly defended the new mode of choosing the executive, claiming among several reasons for the change that "No body had appeared to be satisfied with an appointment by the Legislature" (Farrand 1911b, p. 500; also see pp. 501–502).

The issue was taken up again the next day, September 5, when Charles Pinckney again expressed his objections to the new mode and, among other reasons, stated that "electors will not have sufficient knowledge of the fittest men" (Farrand 1911b, p. 511). John Rutledge said that he was "much opposed to the plan reported by the Committee" and moved "to postpone the Report under consideration & take up the original plan of appointment by the Legislature" (Farrand 1911b, p. 511). George Mason admitted that he had "objections to an appointment by the Legislature as originally planned," but that he also had "objections to the mode proposed by the Committee" (Farrand 1911b, p. 512). Similar to Mason's concern the day before, Williamson said that he "could not agree to the clause without some such modification," as he was also concerned that the new mode would throw the election into the Senate (Farrand 1911b, p. 512). By the end of the day's proceedings, Mason concluded that "As the mode of appointment is now regulated, he could not forbear expressing his opinion that it is utterly inadmissible. He would prefer the government of Prussia to one which will put all power into the hands of seven or eight men [senators], and fix an aristocracy worse than absolute monarchy" (Farrand 1911b, p. 515).

The next day, September 6, several delegates continued to voice opposition to the new mode of selecting the executive, particularly of concern was the belief

that it would too often throw the choice into the Senate. Nonetheless, the whole resolution concerning the appointment of electors as reported by the Committee was passed in the affirmative, 9-2, with only two lower South states, North and South Carolina opposed. Afterward, Richard Dobbs Spaight said that "if the election by Electors is to be crammed down, he would prefer their meeting altogether ... at the seat of the General Government," which was promptly defeated by all states except North Carolina (Farrand 1911b, p. 526). After several minor amendments to different aspects of the operation of what would become known as the Electoral College, it was finally agreed that in the case that no candidate received a majority of the electoral votes that "The House of Representatives shall immediately choose by ballot one of them for President, the members from each State having one vote" (Farrand 1911b, p. 527), which finally settled the selection of the executive.

The positions of the delegates during the late August and early September debates on choosing the executive were similar to those taken earlier: other than James Madison, southerners generally did not support selection of the executive by electors. During this time, no delegates (northern or southern) brought up the issue of a disproportion of northern and southern voters nor mentioned the issue of slaves. In the final analysis, northern (eastern) delegates supported the concept of presidential electors and lower South delegates and a couple other southerners generally opposed their use. In the final convention vote on electors, the only two states to vote against the whole resolution were the two Carolinas. The evidence from the convention thus does not appear to support Finkelman's (2002) argument that the concept of presidential electors had proslavery origins, even though James Madison said so. Given the lower South's adamant opposition to electors, it is reasonable to conclude that the constitutional imposition of electors can be regarded to have been "costly" to the lower South states.

2.3.3 The Fugitive Slave Clause

There may not have been any real disagreement on a fugitive slave clause (Article IV, Section 2, Clause 3) at the convention, because, as Finkelman (1987) notes, the issue already had been settled as a similar clause was included earlier in the Northwest Ordinance of 1787. A brief examination of the convention's consideration of this issue suggests this interpretation appears to be correct.

While there was no explicit discussion of including a fugitive slave provision in the Constitution until the end of August, the issue of providing some constitutional security for the "property" of southern slaveholders was a continuing theme in the convention. In the midst of the convention's discussion of the apportionment issue on July 12, C.C. Pinckney, in response to Gouverneur Morris' arguments that there should be no representation of slaves, said that "property in slaves should not be exposed to danger under a Govt. instituted for the protection of property" (Farrand 1911a, p. 594). Edmund Randolph also maintained that slaveholders required some security for their "property;" to which Morris wanted to know: "If the Southn. States

get the power into their hands, and be joined as they will be with the interior Country they will inevitably bring on a war with Spain for the Mississippi … What security the Northn. & middle States will have agst. this danger" (Farrand 1911a, pp. 603–604). Pierce Butler responded that "The security the Southn. States want is that their negroes may not be taken from them which some gentlemen within or without doors, have a very good mind to do" (Farrand 1911a, p. 605). A week and a half later on July 23, in response to Elbridge Gerry's motion to refer the convention's proceedings to a Committee of Detail to prepare a draft of a constitution, C.C. Pinckney said he "reminded the Convention that if the Committee should fail to insert some security to the Southern States agst. an emancipation of slaves, and taxes on exports, he shd. be bound by duty to his State to vote agst. their report (Farrand 1911b, p. 95).

Then during the middle part of August when the convention began discussing the issue of the slave trade, often implicit in the delegates' debates was the idea of a desire for some type of security for southerners' "property" (Farrand 1911b, pp. 364–365, 369–374, 400, 414–417). But the first explicit mention of a fugitive slave clause at the convention was August 28 when the convention took up the "privileges and immunities" clause of the proposed constitution. C.C. Pinckney was the only delegate to speak when he said he "was not satisfied with it. He seemed to wish some provision should be included in favor of property in slaves" (Farrand 1911b, p. 443). Nonetheless, the clause was approved, 9-1-1, without any other discussion. When the convention took up the extradition clause of the proposed constitution next, Pierce Butler and Charles Pinckney immediately moved "to require fugitive slaves and servants to be delivered up like criminals" (Farrand 1911b, p. 443). James Wilson objected, claiming this would require "the Executive of the State to do it, at public expense" (Farrand 1911b, p. 443). Roger Sherman also objected, suggesting it was not appropriate "in the public seizing and surrendering a slave or servant, than a horse" (Farrand 1911b, p. 443). Butler withdrew the motion, indicating he would wait to make it separate from the extradition clause.

The next day, August 29, the fugitive slave clause was formally proposed by Pierce Butler to be inserted as a separate article in the draft constitution. To wit: "If any person bound to service or labor in any of the U-States shall escape into another State, he or she shall not be discharged from such service or labor in consequence of any regulations subsisting in the state to which they escape, but shall be delivered up to the person justly claiming their service or labor" (Farrand 1911b, p. 453). The motion was unanimously agreed to, 11-0, without any discussion whatsoever.

What is obvious from examining the positions of the delegates on the fugitive slave clause is that northern and southern delegates unanimously allied on the clause and southern delegates (especially, lower South delegates) got exactly as they wished. This constitutional clause can be considered to have been unambiguously "beneficial" to slave interests in the overall South.

2.3.4 The Slave Trade Compromise

Constitutional scholars have long traced the adoption of the Constitution's limitation on congressional interference in the international slave trade (Article I, Section 9, Clause 1), the prohibition on export taxes (Article I, Section 9, Clause 5), and the simple-majority congressional voting rule for navigation acts (commercial regulations) to a political bargain between northern and southern delegates.[9] In particular, John Rutledge (South Carolina) and a few other (unnamed) lower South delegates are said to have agreed to vote against a super-majority voting rule (a two-thirds requirement) to enact commercial regulations, including taxation of trade, in exchange for support from Roger Sherman (Connecticut) and a few other (unnamed) New England delegates for a twenty-year extension of the limitation on the national government's ability to prohibit the international slave trade and for the export tariff prohibition.[10] The bargain essentially eased the process whereby the national government could tax and regulate commerce in exchange for putative restrictions on the scope of its taxation and regulatory powers.

Scholars accord this North-South agreement a *sine qua non* status among the various political bargains that brought the Constitutional Convention to a conclusion.[11] Indeed, Gary Walton and Hugh Rockoff (2002, p. 271) contend that the slave trade clause was "part of one of the great constitutional compromises" to which our "nation's forefathers agreed in 1787." The presumed benefits to slave interests led another scholar to label it the "dirty compromise" of the convention (Finkelman 1987, p. 214).

The overall agreement, which Farrand (1904, 1911a,b) refers to as the "Slave Trade Compromise," plays an important role in this chapter's thesis that the various North-South alliances at the convention were "costly" to the South, particularly to the lower South. Previous accounts of the Compromise overlook the fact that economic principles would dash southern framers' aspirations for the export tax prohibition. Moreover, the simple majority voting rule for commercial legislation would make it easier for these hopes to be dashed. Southern interests ended up with an emasculated prohibition on export taxes and a twenty-year moratorium on a national prohibition on slave importation that lasted no longer than its period of constitutional immunity and which was enacted when almost all states had or soon would have their own prohibitions on slave importation.

A driving force for the Constitutional Convention was that the Articles of Confederation did not give the Continental Congress the power to tax. Congress had

[9] This section borrows freely from Baack et al. (2009).

[10] Because the decisions at the Philadelphia convention were based on the support of state delegations not individual delegates per se and the size of the state delegations ranged from three to eight delegates, the vote of one or two delegates from a particular state delegation could have easily changed that state's vote on particular issues.

[11] See, for examples, Curtis (1858, vol. 2), Farrand (1904), Warren (1928), Jensen (1964), Kelly and Harbison (1970), Finkelman (1987), and McGuire (2003).

to rely on requisitions and voluntary contributions from the states. This institutionalized the free rider problem because individual states had little incentive to make contributions. Unable to authorize imposts (tariffs) during the Revolutionary War, Congress circumvented the free rider problem by printing currency and authorizing the Continental Army to confiscate property. With the realization that these policies were an unsatisfactory way of financing the government, a goal of many convention delegates was to resolve the free rider problem by granting Congress the power of taxation, including the tariffs it had unsuccessfully sought during the war. An important issue at the Convention was whether *export* tariffs would be allowed. Not surprisingly, nearly all southern delegates opposed export tariffs. Allowing taxes on exports was obviously contrary to southern interests in that it could foreshadow a national tax regime in which the southern states would be the government's "cash cow."

The argument that the Slave Trade Compromise was "costly" to the South is grounded in a proposition in international economics. To wit: hiding behind any set of import tariffs is a regime of *de facto* export tariffs. This symmetry is known as the "Lerner Symmetry Theorem," after Lerner (1936). It has obvious implications for any discussion of the constitutional prohibition on explicit exports tariffs.[12] In particular, the Constitution, since it permits import tariffs, in no way prevents *implicit* export tariffs. Moreover, a simple majority voting rule for enacting import tariffs makes these implicit export tariffs easier to enact than would be the case with a super majority voting rule. To be sure, the clause precludes singling out particular exports for tariffs, but whether this prevents import tariffs from having regionally disparate effects depends on the regional composition of exports.

Tariff symmetry is similar to the parallel in the public finance literature between broad-based income and broad-based expenditure taxes, where an equivalent tax rate on income or on spending (including future spending) has identical economic consequences (Browning and Browning 1994, pp. 309–313). Because exports represent a country's foreign earnings and imports its foreign expenditures, a government can tax people when they export (earn) or when they import (spend). Symmetry traces to the fact that a nation's ability to import is ultimately limited by its ability to export. Buying internationally mirrors selling internationally, just as selling mirrors buying. Whether governments tax buying or selling internationally is irrelevant, either tax affects relative prices the same. One raises the price of imported goods; the other lowers the price of exported goods. Thus, either tax increases the incentive to produce import-competing goods *and* decreases the incentive to produce export goods. Import tariffs do this by increasing the profitability of

[12] Awareness of tax symmetry pre-dates Lerner, going back at least to the seventeenth century (Irwin 1996). Irwin considers symmetry one of the three most important propositions in international economics.

domestic production of import-competing goods relative to other goods, including goods for export. Export tariffs decrease the profitability of domestic production of goods for export relative to other goods, including import-competing goods.[13]

2.3.5 *Taxation of Exports and Imports*

Scholars trace the export tax prohibition in the Constitution (Article I, Section 9, Clause 5) to the influence of southern slaveholding delegates.[14] There is consensus among constitutional scholars concerning the "facts" behind the drafting of the export clause: Southern delegates knew that the nation's exports came primarily from their states. Export taxes would reduce southern exports. Constitutionally preventing such taxes, the southern delegates reasoned, would protect their states' economies. Accordingly, southern delegates insisted on an export tax prohibition in the Constitution.[15]

C.C. Pinckney first brought up the issue of taxation of exports on July 12 in response to the views expressed by Gouverneur Morris concerning representation and taxation, in particular Morris's view that direct taxation be in proportion to representation, but "With regard to indirect taxes on exports & imports & on consumption, the rule would be inapplicable" (Farrand 1911a, p. 592). Reacting to Morris's statements, Pinckney said that he was "alarmed at what was said [by Morris] yesterday, concerning Negroes ... was now again alarmed at what has been thrown out concerning the taxing of exports ... hoped a clause would be inserted in the system restraining the Legislature from taxing Exports" (Farrand 1911a, p. 592).[16] Then on July 23, as noted above, in response to the motion to commit the proceedings to the Committee of Detail, C.C. Pinckney "reminded the Convention that if the committee should fail to insert some security to the Southern States agst. ... taxes on exports, he shd. be bound by duty to his State to vote agst. their report" (Farrand 1911b, p. 95).

[13] McGuire and Van Cott (2002) explain tariff symmetry in a straightforward supply and demand setting.

[14] For discussion of the details of southern involvement in the export tariff prohibition, see Curtis (1858, vol. 2), Farrand (1904), Warren (1928), Jensen (1964), Kelly and Harbison (1970), Finkelman (1987), and Mcmillin (2004).

[15] For discussion of these "facts" among traditional historians, see Jensen (1964). For discussion among legal scholars, see Finkelman (1987). For a modern economic examination of the Constitution's design espousing these "facts," see McGuire (2003). I was unable to find any discussion in the constitutional and historical literature that disputes the view that southern slaveholding interests were the driving force behind the export tariff clause.

[16] In discussing the apportionment issue the day before, Gouverneur Morris had said that he could "never agree to give such encouragement to the slave trade as would be given by allowing them a representation for their negroes" (Farrand 1911a, p. 588).

Following the report of the Committee of Detail on August 6, the subject of export taxes was addressed on August 16 when the convention was discussing the power of the national legislature "to lay and collect taxes, duties, imposts and excises" (Farrand 1911b, p. 95). On that day, George Mason "urged the necessity of connecting with the power of levying taxes duties &c, <the prohibition in Sect 4 of art VI [VII]> that no tax should be laid on exports. He was unwilling to trust to its being done in a future article. He hoped the Northn. States did not mean to deny the Southern this security" (Farrand 1911b, p. 305). Mason then moved to insert the following into the tax powers clause: "provided that no tax duty or imposition, shall be laid by the Legislature of the U. States on articles exported from any State" (Farrand 1911b, p. 305). In short, Mason's concern was that even though the Committee report included an identical prohibition on export taxes, albeit three sections later in the report, Mason wanted the prohibition to be included explicitly in the earlier tax powers clause because he did trust it (northerners?) to be done later! The discussion that ensued was quite contentious.

Interestingly, Roger Sherman, the reputed broker of the Slave Trade Compromise for the New England states, had no objections other than that the motion would unbalance the reported order. (Oliver Ellsworth, Sherman's fellow Connecticut delegate, was likewise against taxing exports, but also did not want to unbalance the order as reported (Farrand 1911b, p. 307).) John Rutledge, reputedly the broker of lower South participation in the Compromise, didn't care where the prohibition was put as he would "vote for the clause as it stood" (Farrand 1911b, p. 306). But Gouverneur Morris considered it "as inadmissible any where. It was so radically objectionable, that it might cost the whole system the support of some members" (Farrand 1911b, p. 306). According to Morris, "it would not in some cases be equitable to tax imports without taxing exports . . . taxes on exports would be often the most easy and proper of the two" (Farrand 1911b, p. 306). James Madison also supported the use of export tariffs, arguing that taxing exports "might with particular advantage be exercised with regard to articles in which America was not rivalled in foreign markets, as Tobo. &c." and thus the tax could raise the price of the articles and consequently have the foreign consumer pay it (Farrand 1911b, p. 306). While James Wilson said he was "decidedly agst prohibiting general taxes on exports," John Mercer (Maryland) maintained he was "strenuous against giving Congress power to tax exports. Such taxes were impolitic" (Farrand 1911b, p. 307). Elbridge Gerry apparently favored the prohibition on taxing exports, as he said "the legislature could not be trusted with such power. It might ruin the country. It might exercise partially, raising one and depressing another part" (Farrand 1911b, p. 307). In response, Gouverneur Morris argued that "[H]owever the legislative power may be formed, it will if disposed be able to ruin the Country . . . Taxes on exports are a necessary source of revenue" (Farrand 1911b, p. 307). Finally, the delegates agreed that the issue of export taxes should lie over for the place where it was reported.

The convention then took up the export tax issue on August 21, when the split along sectional lines on export tariffs continued. Pierce Butler made known that he "was strenuously opposed to a power over exports; as unjust and alarming to the staple States" (Farrand 1911b, p. 360). Gouverneur Morris, though, believed that

"Tobacco, lumber, and live-stock are three objects belonging to different States, of which great advantage might be maed by a power to tax exports ... by which a tax might be thrown on other nations" (Farrand 1911b, p. 360). "The power of taxing exports may be inconvenient at present;" according to John Dickinson (Delaware), "but it must be of dangerous consequence to prohibit it with respect to all articles and for ever" (Farrand 1911b, p. 361). James Wilson also favored "the general power over exports," while Elbridge Gerry said he was "strenuously opposed to the power over exports" (Farrand 1911b, p. 362). But George Mason argued that "[I]f we compare the States in this point of view the 8 Northern States have an interest different from the five Southn. States ... The Southern States had therefore ground for their suspicions. The case of exports was not the same with that of imports. The latter were the same throughout the States: the former very different. As to Tobacco other nations do raise it ... The impolicy of taxing that article had been demonstrated by the experiment of Virginia" (Farrand 1911b, p. 362–363).

Just before the final vote on the issue, James Madison argued that "In order to require 2/3 of each House to tax exports – as a lesser evil than a total prohibition <moved to insert the words 'unless by consent of two third of the Legislature'>," which was seconded by James Wilson, but passed in the negative, 5-6, with all five southern states voting no. Madison recorded the Virginia votes, indicating that he and George Washington voted in favor of the motion (Farrand 1911b, p. 363). When the convention then voted on the export clause, it passed in the affirmative, 7-4, with all five of the southern states voting yes, but with George Washington and James Madison having voted no (Farrand 1911b, p. 363–364).[17] The next day, responding to Gouverneur Morris's wish to have the export tax prohibition included with a commitment to a committee of all states to consider the issue of slave importation, slave taxation, and navigation acts, Pierce Butler declared that he "never would agree to the power of taxing exports" (Farrand 1911b, p. 374). The prohibition on export taxes was not committed and remained in the Constitution.

That the desires of southerners for a prohibition on export tariffs would be defeated, not at the convention, but by the national government's ability to enact *import* tariffs (Article I, Section 8, Clause 1) raises an important question. Namely, what was the posture of the southern delegates on national import tariffs? Southern delegates were suspicious of commercial regulations by the national government and sought to make such regulation "costly" to enact, as support by many south-

[17] McGuire (2003, pp. 70–72) estimates that delegates who represented states with the greatest concentration of slaves, *ceteris paribus*, were much more likely to vote in favor of prohibiting export tariffs. It should be noted, however, that because the individual delegate votes in McGuire (2003) are inferred from, among other information, the recorded voting sentiments of the convention delegates, the votes are not independent information that can then be used to corroborate whether a particular delegate voted as part of a specific North-South alliance at the Constitutional Convention. In other words, the delegate votes cannot be used as a dependent variable in a regression model to test whether a specific delegate participated in a particular North-South agreement or compromise because the votes are inferred from the delegates' sentiments expressing how they were going to vote.

erners for a two-thirds congressional voting rule for commercial regulations would indicate. Yet because they understood the importance of tariff revenue for financing government expenditures, southern delegates did not oppose actively import tariffs at the convention. In fact when the basis of the first part of Article I, Section 8, Clause 1 ("Congress shall have Power To lay and collect Taxes, Duties, Imposts and Excises") was first introduced into the convention as the second resolution of the New Jersey Plan, there were no specific objections to a power to tax.

However, according to the notes of James McHenry (Maryland), the Maryland delegation was not too pleased with the possible consequences of the tax power clause when the convention first took up the Committee of Detail's draft constitution, which contained the tax clause from the New Jersey Plan. As McHenry said about his conversation on August 7 with his fellow Maryland delegate, Daniel Carroll, about the clause, "[w]e almost shuddered at the fate of the commerce of Maryland should we be unable to make any change in this extraordinary power. We agreed that our deputation ought never to assent to this article in its present form" (Farrand 1911b, pp. 211–212). McHenry noted that two other Maryland delegates, Daniel of St. Thomas Jenifer and John Mercer, entered into the conversation later and generally agreed with his and Carroll's sentiments.

When the tax clause was taken up on August 16, Daniel Carroll "reminded the Convention of the great difference of interests among the States," as the two issues discussed were (1) the difference between imposts (taxes on commerce) and duties (taxes on most any object) and (2) George Mason's desire to include the export tax prohibition directly in the tax power clause, as noted above (Farrand 1911b, p. 305–308). To repeat, what nearly all southern delegates wanted was an assurance that exports would not be taxed. With no objections to the general government's power to tax per se, the tax power clause then passed in the affirmative nearly unanimously with Elbridge "Gerry alone answering no" (Farrand 1911b, p. 308).[18]

What can be concluded about the positions of the delegates on the export tax prohibition? First, southern delegates were nearly unanimous in their support for a prohibition on export taxes; especially vocal were George Mason and the lower South delegates. Two *upper* South delegates, James Madison and George Washington, opposed the prohibition, as they supported export taxes. Third, most northern (eastern) delegates, with the notable exception of the Connecticut delegation, adamantly opposed the prohibition, especially vocal was Gouverneur Morris. Fourth, a couple of New England delegates (Oliver Ellsworth and Roger Sherman) supported strongly the southern position on the prohibition during the debates.

[18]There were, though, various changes to the remainder of Article I, Section 8, Clause 1, later in the convention. These changes amended and/or clarified the purposes of government tax revenues and inserted the uniformity limitation on duties, excises, and imposts, but no changes were made to the "power to lay and collect taxes" itself (Farrand 1911b, pp. 326–327, 355–356, 392, 412–414, 418, 470, 481, 497, 499, 529–530, 614).

As a result, the southern delegates (other than Madison and Washington) were successful in getting the prohibition in the Constitution, but gained little or nothing because of tariff symmetry.

2.3.6 Slave Importation

The constitutional restriction on national interference in the international slave trade, a twenty-year (rather than a shorter or no) restriction on a national prohibition on slave importation (Article I, Section 9, Clause 1), would appear to be an obvious "benefit" to the South.[19] But its inclusion in the Constitution raises some interesting questions, because prior to or shortly after the Constitution became effective nearly all states had legally banned slave importation. "Thus," according to Phillips (1918 [1966], p. 183), "at the time when the framers of the Federal Constitution were stopping congressional action for twenty years, the trade was legitimate only in a few of the Northern states, all of which soon enacted prohibitions, and in Georgia alone in the South." Georgia prohibited slave imports from the West Indies, Bahamas, and Florida in 1793 and banned imports from African in 1798 (Phillips 1918 [1966], p. 133). South Carolina repealed its prohibition effective January 1, 1804. Slaves were thereafter imported into South Carolina in large numbers until a national prohibition became effective January 1, 1808.[20]

Many southern delegates, as noted earlier, declared during July that their states had to have security for their slaves. Likely in response, the Committee of Detail included in its draft constitution, which was reported on August 6, the slave trade clause immediately following the export tax clause, "No tax or duty shall be laid by the Legislature on articles exported from any State." The new clause then stated, "nor on the migration or importation of such persons as the several States shall think proper to admit; nor shall such migration or importation be prohibited" (Farrand 1911b, p. 183). Two days later on August 8 while the convention was discussing a section on future representation (due to a change in a state's size or because of new states) in the reported draft, Rufus King, who had long been opposed to "the admission of slaves into the rule of Representation" but not vocal about it, was now quite concerned given the tenor of the report, as he claimed, "In two great points the hands of the Legislature were absolutely tied. The importation of slaves could not be prohibited – exports could not be taxed. Is this reasonable? ... If slaves are to be imported shall not the exports produced by their labor, supply a revenue the better

[19]The slave trade clause did allow a tax or duty up to ten dollars per slave, but no tariff on slave imports was ever enacted.

[20]Details of state restrictions on slavery and the slave trade can be found in Phillips (1918 [1966], pp. 132–133) and Fogel and Engerman (1974, pp. 33–34). Du Bois (1896 [1969], pp. 223–239) presents somewhat different information on slave trade restrictions for Massachusetts, New Hampshire, and New York.

to enable the Genl. Gout. to defend [in case of invasion or sedition] their Masters?" (Farrand 1911b, p. 220). Although the convention had not yet taken up the slave trade clause, the debate on the trade had begun.

King insisted that there must at least be some time limitation on slave imports, or no representation for them or taxes on exports. Gouverneur Morris was as adamant in his opposition to slave representation, no limits on the slave trade, and the prohibition on export taxes, and so he moved to insert the word "free" in front of "inhabitants" in the section under consideration so that future representation would be based on the number of "free inhabitants" only. In his view, the draft constitution allowed southern slaveholders to free ride on the sacrifices of northerners (easterners), because "the Southern States are not to be restrained from importing fresh supplies of wretched Africans, at once to increase the danger of attack, and the difficulty of defense; nay they are to be encourage to it by an assurance of having their votes in the Natl Govt increased in proportion. and (sic) at the same time to have their exports & their slaves exempt from all contributions for the public service" (Farrand 1911b, pp. 222–223). Morris concluded that he did not want to "saddle posterity with such a Constitution" (Farrand 1911b, p. 223). In response, Roger Sherman did not believe the issue of slaves and representation was "liable to such insuperable objections," and Charles Pinckney said that he "considered the fisheries & the Western frontier as more burdensome to the U. S. than the slaves" (Farrand 1911b, p. 223). James Wilson "thought the motion premature," after which Morris's motion was promptly defeated, 1-10 (Farrand 1911b, p. 223).

The convention did not debate the slave trade clause again until August 21 when, following agreement on the export tax prohibition, they were taken up. Luther Martin, an *upper* South delegate, immediately proposed to prohibit slave imports or to tax them to prevent importation, because "such a clause wd. leave an encouragement to this trafic . . . slaves weakened one part of the Union which the other parts were bound to protect . . . it was inconsistent with the principles of the revolution and dishonorable to the American character to have such a feature in the Constitution" (Farrand 1911b, p. 364). John Rutledge replied that "[R]eligion & humanity had nothing to do with this question—Interest alone is the governing principle with Nations—The true question at present is whether the Southn. States shall or shall not be parties to the Union. If the Northern States consult their interest, they will not oppose the increase of Slaves which will increase the commodities of which they will become the carriers" (Farrand 1911b, p. 364). "[L]et every State import what it pleases," exclaimed Oliver Ellsworth, "[t]he morality or wisdom of slavery are considerations belonging to the States themselves" (Farrand 1911b, p. 364). Charles Pinckney then proclaimed "South Carolina can never receive the plan [Constitution] if it prohibits the slave trade" (Farrand 1911b, p. 364).

The next day, Roger Sherman said he "was for leaving the clause as it stands" even though he "disapproved of the slave trade: yet as the States were now possessed of the right to import slaves . . . it best to leave the matter as we find it" (Farrand 1911b, p. 369). George Mason, another *upper* South delegate, expressed his *opposition* to the slave trade, which he considered "nefarious traffic. As to the States being in possession of the Right to import, this was the case with many

other rights, now to be properly given up ... the Genl. Govt. should have power to prevent the increase of slavery" (Farrand 1911b, p. 370). Oliver Ellsworth believed a prohibition on the trade would "be unjust towards S. Carolina & Georgia," while Charles Pinckney said that "[A]n attempt to take away the right as proposed will produce serious objections to the Constitution" (Farrand 1911b, p. 371). C.C. Pinckney insisted that "S. Carolina & Georgia cannot do without slaves. As to Virginia she will gain by stopping the importations. Her slaves will rise in value, & she has more than she wants. It would be unequal to require S.C. & Georgia to confederate on such Unequal terms" (Farrand 1911b, p. 371).[21] He continued, admitting "it to be reasonable that slaves should be dutied like other imports, but should consider a rejection of the clause as an exclusion of S. Carols from the Union" (Farrand 1911b, pp. 371–372).

The debates on the slave trade on August 22 did not let up, as one after another delegate expressed his position on the trade (Farrand 1911b, pp. 372–373). Abraham Baldwin (Georgia) and Hugh Williamson were both adamant that it was a local matter and should be left up to the states to decide. In one form or another, John Dickinson, Elbridge Gerry, Rufus King, John Langdon (New Hampshire), James Madison, and James Wilson all were likewise adamant that the trade was a national matter and should be subject to general government, not state government, control, if not subject to an outright prohibition. Recognizing the impasse, C.C. Pinckney made a motion that he hoped would remove one of the problems: to wit "to commit [to a committee of the states] the clause that slaves might be made liable to an equal tax with other imports," and John Rutledge, who said he was "strenuous agst. striking out the Section," seconded the motion (Farrand 1911b, p. 373). Gouverneur Morris responded by laying out explicitly the components of the Slave Trade Compromise, wishing that "the whole subject to be committed including the clauses relating to taxes on exports & to a navigation act. These things may form a bargain among the Northern & Southern States" (Farrand 1911b, p. 374). Reacting strongly, Pierce Butler said, as noted earlier, that he "never would agree to the power of taxing exports," while Roger Sherman declared, "it was better to let the S. States import slaves than to part with them, if they made that a *sine qua non*" (Farrand 1911b, p. 374). Sherman then explained that as the export tax prohibition had already been agreed to it "therefore could not be committed" (Farrand 1911b, p. 374).

After several delegates spoke in favor of committing the clauses in an attempt to resolve the impasse, Oliver Ellsworth, in a very prescient moment, said he was "for taking the plan as it is. This widening of opinions has a threatening aspect. If we do not agree on this middle & moderate ground ... [the states] should fly into a variety of shapes & directions, and most probably into several confederations and not without bloodshed" (Farrand 1911b, p. 375). The convention, of course, did

[21] Interestingly, just prior to the start of the Constitutional Convention in spring 1787, the same South Carolina delegates (actually, three of the four; Pierce Butler, Charles Pinckney, and John Rutledge) voted in the South Carolina state legislature to prohibit slave imports into their state (Brady 1972, p. 602).

not accept "this middle & moderate ground," as it agreed to commit the clauses. When the committee reported to the convention on August 24, the clauses under consideration had been amended, indicating (1) that the national legislature could not prohibit the international slave trade prior to 1800, (2) that a tax on slave imports up to the average duties on other imports could be levied, and (3) the two-thirds congressional vote requirement to enact navigation acts was to be struck out of the proposed constitution (Farrand 1911b, p. 400).

The convention took up the committee report on the clauses the next day. C.C. Pinckney made the first motion which was to strike out "the year eighteen hundred" and insert "the year eighteen hundred and eight," meaning the national legislature could not prohibit the international slave trade prior to 1808, which passed, 7-4, with the support of all southern states except Virginia, and New Jersey, Pennsylvania, and Delaware also voted no (Farrand 1911b, p. 415). Only James Madison spoke on the motion, objecting that "all the mischief that can be apprehended from the liberty to import slaves" will be produced in the twenty years of the prohibition (Farrand 1911b, p. 415). After amending the wording of the clause in an attempt to limit the prohibition on national legislative action to those states that had not yet prohibited slave imports themselves ("importation of slaves into such of the states as shall permit the same shall not be prohibited"), the entire amended clause passed in the affirmative, 7-4, with support of all southern states except Virginia. (New Jersey, Pennsylvania, and Delaware again also voted no.) Then after some discussion on whether a duty on slaves meant the Constitution was "acknowledging men to be property," the convention unanimously agreed to amend the slave duty clause to allow "a tax or duty .., on such importation not exceeding ten dollars for each person" (Farrand 1911b, pp. 416–417). (The clause on the congressional voting requirement for navigation acts was immediately thereafter postponed.)

What can be said about the positions of the delegates on the slave trade clause in the Constitution? First, the southern delegates were divided on slave importation. While the *lower* South delegates (in particular, Charles and C.C. Pinckney and John Rutledge) were staunch supporters of continuing to allow states to import slaves if they wish, the *upper* South delegates (in particular, James Madison, Luther Martin, and George Mason) were equally as staunch in their opposition to the slave trade. Second, a couple New England delegates (in particular, Oliver Ellsworth and Roger Sherman) supported the *lower* South position throughout the debates, often arguing explicitly that it would not be fair if Georgia and the Carolinas could not import slaves as they saw fit. Third, most northern (eastern) delegates, other than the Connecticut delegation, adamantly opposed the slave trade. Fourth, while the *lower* South delegates initially opposed a duty on slave imports, they compromised on the issue and eventually proposed allowing the possibility of a duty themselves, in an attempt to forestall any movement to ban the slave trade outright. Fifth, the northern (eastern) delegates appear to have wanted a tax on slave imports throughout the entire debates on the issue.

Given the nature of the economies of the two regions in the South, in which the *upper* South had a surplus of slaves and was thus an exporter of slaves to other regions and in which the *lower* South had an expanding plantation economy and

was consequently an importer of slaves, slave importation would appear to have been beneficial to the *lower* South, but potentially harmful (through lower slave prices) to the *upper* South.

2.3.7 Congressional Voting Rules for Commercial Regulations

C.C. Pinckney's reminder to his fellow delegates on July 23 that the southern states desired security against emancipation of their slaves and against export taxes might have served as a catalyst for the Committee of Detail's decision to include in its report a congressional obstacle to enacting navigation acts. The Committee likely recognized that southern delegates were concerned about northern dominance and commercial regulations, maybe Pinckney reminded them of this. Regardless of whether his reminder was a catalyst, the Committee's report on August 6 included a clause stating, "No navigation act shall be passed without the assent of two thirds of the members present in each House" (Farrand 1911b, p. 183).

There was no discussion or mention of the two thirds rule until, as noted above, Gouverneur Morris, during the debates on the slave trade clause on August 22, laid out the elements of the Slave Trade Compromise, suggesting that the entire subject of the slave trade, export taxes, and voting on a navigation act should be committed. To repeat, Morris believed "[t]hese things may form a bargain among the Northern & Southern States" (Farrand 1911b, p. 374). The commitment was agreed to, the committee reported back 2 days later with a proposal to strike out the navigation act clause, the committee report was taken up the next day (August 25), and consideration of the navigation act clause was postponed.

The convention did not take up the clause for several days; when on August 29, Charles Pinckney moved to delay consideration of the committee's report (to strike out the two-thirds vote requirement) in favor of a new clause. To wit: "That no act of the Legislature for the purpose of regulating the commerce of the U–S. with foreign powers, or among the several States, shall be passed without the consent of two thirds of the members of each House" (Farrand 1911b, p. 449). Pinckney argued that the two-thirds motion be considered because "different interests would be a source of oppressive regulations if no check to a bare majority should be provided" (Farrand 1911b, p. 449). He further maintained, "[t]he power of regulating commerce was a pure concession on the part of the S. States" (Farrand 1911b, p. 449). His cousin, C.C. Pinckney, likewise argued that "it was the true interest of the S. States to have no regulation of commerce," but citing the "liberal conduct toward the views of South Carolina, and the interest the weak Southn. States had in being united with the strong Eastern States," it was only proper that "no fetters should be imposed on the power of making commercial regulations" (Farrand 1911b, pp. 449–450). George Clymer (Pennsylvania) agreed that there should be no "unnecessary restrictions. The Northern & middle States will be ruined, if not enabled to defend themselves against foreign regulations" (Farrand 1911b, p. 450). Roger Sherman believed that

"to require more than a majority to decide a question was always embarrassing," while Gouverneur Morris considered "the motion as highly injurious" (Farrand 1911b, p. 450).

The debates, which were quite contentious, continued for some time as one after another delegate took a position on this important issue. Hugh Williamson said that he favored "making two thirds instead of a *majority* requisite, as more satisfactory to the Southern people" (Farrand 1911b, p. 450; emphasis in original). George Mason, venting his frustration over the lack of support for the two-thirds requirement, maintained that "[t]he Majority will be governed by their interests. The Southern States are the *minority* in both Houses. Is it to be expected that they will deliver themselves bound hand & foot to the Eastern States, and enable them to exclaim, in the words of Cromwell on a certain occasion – 'the lord hath delivered them into our hands"' (Farrand 1911b, p. 451; emphasis in original). Likewise, in Edmund Randolph's view "there were features so odious in the Constitution as it now stands . . . A rejection of the motion would compleat the deformity of the system" (Farrand 1911b, p. 452).

On the other side of the debate were Pierce Butler, Nathaniel Gorham, James Madison, John Rutledge, Richard Dobbs Spaight, and James Wilson, who all opposed the motion. Given their reputed participation in the Slave Trade Compromise, the positions of Butler and Rutledge are of particular interest. Butler "differed from those who considered the rejection of the motion as no concession on the part of the S. States. He considered the interests of these and of the Eastern States, to be as different as the interests of Russia and Turkey. Being notwithstanding desirous of conciliating the affections of the East: States, he should vote agst. requiring 2/3 instead of a majority" (Farrand 1911b, p. 451). Similarly, John Rutledge was against the two thirds vote, as he believed that a majority requirement would not be abused. "At the worst a navigation act could bear hard a little while only on the S. States. As we are laying the foundation for a great empire, we ought to take a permanent view of the subject and not look at the present moment only" (Farrand 1911b, p. 452). So, Butler acknowledged the radically different interests between the two regions, but would vote against requiring a two thirds vote for commercial regulations anyway, likely as conciliation with the northern (eastern) states. And Rutledge believed that, even if a simple-majority vote for commercial regulations "cost" the South in the short run, it was in the long run interest of the nation.

The final vote to consider the two-thirds requirement for commercial regulations, with foreign nations as well as among the states, passed in the negative, 4-7, with all southern states except South Carolina voting yes (Farrand 1911b, p. 553).[22] The

[22]Despite the possibility of obscuring effects of vote trading on this integral part of the bargain, McGuire (2003, pp. 77–79) indicates that delegates with merchant interests (mostly from New England) unanimously opposed the proposal and southern delegates from states with greater concentrations of slaves, *ceteris paribus*, were statistically more likely to have supported it. His results also hint at the possibility that delegates who personally owned slaves might have been *less* likely to vote yes. Consequently, the effect on the vote of personal slaveholdings might have been

committee report to strike out the section in the draft constitution "requiring two third of each House to pass a navigation act was then agreed to" unanimously.

Near the end of the convention several delegates made known their objections to the proposed constitution. One of them was Edmund Randolph, who on September 10 stated the most serious objections to the new system of government and included among them "the want of some particular restraint on Navigation acts" (Farrand 1911b, p. 631). As late as September 15, just 2 days before the convention finished business, George Mason was so discontented "at the power given to Congress by a bare majority to pass navigation acts" that he made a last ditch effort to amend the proposed Constitution (Farrand 1911b, p. 631). Mason was certain that the failure to adopt a two-thirds congressional vote on regulating commerce "would not only enhance the freight ... but would enable a few rich merchants in Philada, N. York & Boston, to monopolize the Staples of the Southern States & reduce their value perhaps 50 Per Ct" (Farrand 1911b, p. 631).

Apparently, Mason was aware enough of the importance of a two-thirds vote to southern economic interests that he moved to insert the words: "that no law in nature of a navigation act be passed before the year 1808, without the consent of 2/3 of each branch of the Legislature (Farrand 1911b, p. 631). The motion was promptly defeated without discussion, 3-7, with all southern states that were present voting in favor of the motion, except South Carolina; North Carolina was absent (Farrand 1911b, p. 631). Charles Pinckney was concerned about "[t]hese declarations from members so respectable at the close of this important scene," because he too was not without objections, especially "the power of a majority only of Congs over Commerce. But apprehending the danger of a general confusion, and an ultimate decision by the Sword," he would support the Constitution (Farrand 1911b, p. 632).[23]

What can be said about the positions of the delegates on the two thirds vote for commercial regulations? First, most southern delegates, other than those involved in the Slave Trade Compromise, adamantly supported the two thirds requirement; especially vocal were George Mason, Charles Pinckney, and Edmund Randolph. A couple or so southern delegates opposed the two thirds vote, but among

---

opposite the effect on the vote of the amount of slaves in a delegate's state. Perhaps these possible opposite effects are due to the obscuring effects of vote trades on the issue.

[23] George Mason, in a lengthy declaration of his objections to the Constitution that he circulated to several people, went on at even greater length about the voting rule for navigation acts. To wit: "by requiring only a majority to make all commercial and navigation laws, the five Southern States, whose produce and circumstances are totally different from that of the eight Northern and Eastern States, may <will> be ruined, for such rigid and premature regulations may be made as will enable the merchants of the Northern and Eastern States not only to demand an exorbitant freight, but to monopolize the purchase of the commodities at their own price, for many years, to the great injury of the landed interests, and <the> impoverishment of the people; and the danger is the greater as the gain on one side will be in proportion to the loss on the other. Whereas requiring two-thirds of the members present in both Houses would have produced mutual moderation, promoted the general interest, and removed an insuperable objection to the adoption of this <the> government" (Farrand 1911b, pp. 639–640).

southern states only South Carolina voted against considering the two thirds vote requirement. Third, most northern (eastern) delegates steadfastly opposed the super-majority vote requirement; especially vocal were Gouverneur Morris and Roger Sherman. Consequently, delegates from the northern states and South Carolina were successful in defeating a two thirds requirement for commercial regulations, which in the long run "cost" overall southern interests dearly.

2.3.8 Analysis of the Slave Trade Compromise

Why would *lower* South delegates at the Constitutional Convention agree to this Compromise? In terms of the slave trade clause, the obvious answer is that despite slave trade restrictions in effect or just around the corner in most states they wanted a prohibition on *national* government interference in the slave trade. The trade was important to Georgia and South Carolina, as both were importing slaves from the end of the Revolutionary War to at least the time of the Constitutional Convention (Goldfarb 1994, pp. 22–23; Mcmillin 2004, pp. 30–48). Moreover, Georgia and South Carolina kept the trade open during some of the post-War years when their state restrictions were not in effect.[24]

But how important were slave imports to the rest of the South? Apparently, not important at all! Although slaves were imported in large numbers during the post-War period, all were imported into Georgia and North and South Carolina, as Maryland and Virginia were sending their surplus slaves to other states during the period. Mcmillin (2004, pp. 30–48, Tables 7 and 9) presents a detailed breakdown by year and by state of entry of slave imports for 1783–1810, estimating that South Carolina imported 97,900 slaves, Georgia imported 32,200 slaves, North Carolina imported 500 slaves, and Maryland and Virginia imported zero slaves. Another 39,700 slaves were imported into Florida, Mississippi, and Louisiana, which were Spanish and French territory during much of the post-War period. Accordingly, McMillin estimates that 130,600 of the 170,300 North American slave imports were imported by those states involved in drafting the Constitution, with all but 500 imported into Georgia and South Carolina.[25]

[24] Even when in effect, the state restrictions are commonly believed to have been porous. There is little reason to believe that a national restriction on slave imports would not have been porous as well. In fact, it has been suggested that as many as a quarter million slaves were smuggled into the country before 1860 even though a national prohibition on imports took effect in 1808 (Walton and Rockoff 2002, p. 271).

[25] Fogel and Engerman (1974, pp. 24–25) provide an often-cited indirect estimate of 291,000 slave imports into "North America" for 1780–1810, but their estimate does not include any year or state-by-state breakdown.

As these figures indicate, the lower South gained from extending the restriction on national government prohibition of the trade.[26] But the lower South framers gave up the two-thirds congressional vote to get the twenty-year moratorium (and the prohibition on export tariffs). And a super-majority vote for import tariff legislation would have gone a long way toward closing the "constitutional tax loophole" hiding behind the export tariff prohibition.

So to repeat, why would some *lower* South delegates at the convention participate in the Compromise? Given the export driven nature of the southern economy, perhaps they were so focused on export issues that they overlooked tariff symmetry. Perhaps they were unaware of symmetry. Perhaps they wanted to prohibit taxation of particular exports even if exports in general were still subject to taxation. Perhaps they simply made a mistake. Short of a time machine *and* a truth detector, we can never "know" for certain why the framers made the decisions they did. What we can say, though, is that the willingness of a few lower South framers to trade the two-thirds vote to get the export tariff prohibition and the slave trade extension meant the *overall* South paid a steep price for a twenty-year restriction on a *national* prohibition on the slave trade. Apparently, unintended economic and political consequences of the decisions of the lower South participants in the Compromise trumped their constitutional intentions, as the bargain appears to have been quite costly to the overall southern economy.[27]

2.4 Conclusion

The U.S. Constitution was the outcome of a bargaining process among delegates representing various ideological, sectional, and economic interests. What if economic and political principles ended up trumping a brokered set of provisions in

[26] Finkelman (1987, p. 221) contends that the constitutional ban on national interference in the international slave trade could not have been much of a "lure" to southerners for constitutional settlement "because at the time of the convention none of these states was actively importing slaves from Africa". The accuracy of his statement is dubious as Georgia and South Carolina *were* actively importing slaves from the end of Revolutionary War until about the time of the Philadelphia convention (Goldfarb 1994, pp. 22–23; Mcmillin 2004, pp. 30–48, Tables 7 and 9). In Finkleman's view, the key ingredients of what he calls the "dirty compromise" were the passage of the export tariff prohibition to "lure" lower South delegates to defeat passage of the two-thirds super-majority requirement for all commercial regulations (Finkelman 1987, pp. 213–223). This, though, ignores that tariff symmetry blunted the "lure" of the export tariff prohibition.

[27] Similarly, Brady (1972) and Goldfarb (1994) both maintain that lower South framers participated in the bargain because of the expected benefits to their states' dominant economic interests in keeping the slave trade open, even if for only twenty years. Yet neither recognizes that tariff symmetry and the simple-majority vote component of the bargain cost the southern economy. As a result, the extension of the limitation on national interference in the slave trade apparently benefited two *lower* South states, but the other parts of the Slave Trade Compromise cost the overall southern economy, both the upper *and* the lower South.

an otherwise successful constitution? Moreover, what if the interests of those who were instrumental in seeking many of the provisions bore a substantial burden as a consequence? And what if decades of constitutional and fiscal conflict ensued as a result? Issues surrounding these questions have organized the present discussion of the major North-South alliances during the drafting of the U.S. Constitution.

Heretofore, constitutional scholars have focused on the Founding Fathers' aspirations for the various North-South agreements and compromises at the convention. Has this focus been an empty exercise? No. The framers' aspirations on these and many other issues figured importantly in the U.S. Constitution's design and ratification. Nevertheless, an important parallel story lies behind the aspirations for the constitutional provisions that were the result of the North-South agreements, bargains, and compromises. Indeed, other factors would doom southern aspirations for these agreements before the ink was even dry on them.

First, the compromise among northern and southern delegates to count slaves as three-fifths of other persons for purposes of representation and direct taxation never satisfied the most vocal extremes. Many Massachusetts and Pennsylvania delegates—the most vocal being Gouverneur Morris—opposed any representation for the slaves throughout the proceedings. The lower South delegates—the most vocal being C.C. Pinckney and Pierce Butler—insisted on blacks being equal to whites in the ratio of representation throughout the convention. The upper South delegates, who insisted on slave representation and a periodic census, rating slaves at three-fifths, and some northerners (easterners), in particular the Connecticut delegates, who supported slave representation, counting them at least three-fifths, were apparently satisfied.

Second, as far as the election of the executive is concerned, there is little evidence that it was a proslavery compromise. And, even if one considered it as such, the debates make clear that except for James Madison southern delegates at the convention did not show much interest in the concept of "electors" for selecting the president. Delegates from the lower South states of Georgia and South Carolina in fact actively opposed presidential electors.

Third, while most constitutional scholars do not consider the fugitive slave clause, mandating the return of runaway slaves, to be part of any major North-South bargain, one prominent scholar implies that it might have been, as he argues that the clause had proslavery origins (Finkelman 2002). Whether part of a bargain or not, the clause would have been clearly "beneficial" to southern slave interests at the time the Constitution was written.

Fourth, the constitutional prohibition on export tariffs, arguably the centerpiece of the Slave Trade Compromise for southern economic interests, was a triumph of form over substance. The Constitution's simultaneous authorization of import tariffs hid an unwritten mandate for *implicit* export tariffs. Admittedly, this *de facto* export tariff regime was tantamount to an across-the-board tariff system. However, the importance of southern exports in the nation, the importance of tariff revenue for the national government, and the export-driven nature of the southern economy, meant that the financing of government expenditures was sure to bear down on the southern states, particularly in the lower South.

Fifth, when the lower South framers who participated in the Slave Trade Compromise abandoned the super-majority congressional vote for commercial and tax legislation, this exacerbated the Constitution's between-the-lines approval of *de facto* export tariffs. Easier-to-enact import tariffs make a *de facto* export tariff regime easier to enact. Concerning southern export interests, the best that can be said is that taxes on specific exports were constitutionally precluded. But a general *de facto* tax on southern exports, which consisted primarily of a handful of commodities (tobacco, rice, and indigo, and later cotton), was still constitutional. Moreover, it was easier to enact such a tax.

Sixth, the slave trade component of the Slave Trade Compromise, considered a crucial "part of one of the great constitutional compromises" (Walton and Rockoff 2002, p. 271), lacked substantive benefits for *overall* southern economic interests. What was gained was a twenty-year restriction on *national* prohibition of the international slave trade. The slave trade data indicate that only two lower South economies gained from the twenty-year extension of the slave trade.

The bottom line is that the overall South, both the lower and upper South, bore "costs" from the various agreements, bargains, and compromises that were made during the drafting of the Constitution. Moreover, the evidence from the issue positions of their delegates indicates strongly that the lower South states bore even higher "costs" because of their participation in the numerous alliances with northern (eastern) delegates at the convention.

The influence of these North-South alliances during the Constitution's drafting on the nation's historical path *before* the Civil War was not unimportant. The direct aftermath was a seventy-year *de facto* export tax regime and decades of ongoing North-South constitutional and fiscal conflict during which southern opposition to protective import tariffs flourished from the mid-1820s until southern secession in 1861. The culmination of the North-South tariff struggles and sectional conflict was U.S. House passage of the Morrill tariff in May 1860 that noticeably increased import duties—duties that noticeably *decreased* the profitability of southern exports and the value of southern slaves—and eventually southern secession and the Civil War in 1861. As Oliver Ellsworth argued prophetically in 1787 about the slave trade clause, "If we do not agree on this middle & moderate ground ... [the states] should fly into a variety of shapes & directions, and most probably into several confederations and not without bloodshed" (Farrand 1911b, p. 375). His prophecy just took seventy-five years to come true.

The influence of the alliances on the nation's historical path *after* the Civil War also was not unimportant. An important aspect of the aftermath of the North's victory over the South in the Civil War was the nation's "high tariff" era when tariff rates (implicit exports taxes) on dutiable goods remained highly protective at between 40% and 50%, at least in part because of the simple-majority congressional voting rule, until the constitutional sanctioning of the income tax in 1913. Put bluntly, more than a century of American public finance and many decades of ongoing North-South animosities can be understood as a direct consequence of the many agreements, bargains, and compromises made during the drafting of the Constitution.

The implication from our nation's constitutional founding for designing constitutions in the contemporary world is rather pessimistic, particularly for societies with widely divergent interests and beliefs. The divergent factions that inevitably arise in such circumstances mean that the bargains and deals that are necessary to achieve contemporaneous (short run) constitutional settlement will: (1) ultimately produce unforeseen and unintended consequences that are likely to "doom" the original constitution in the long run, if not sooner; or (2) common agreement and compromise and, as a result, constitutional settlement, will not be possible even in the short run.

References

Baack B, McGuire RA, Van Cott TN (2009) Constitutional agreement during the drafting of the constitution: a new interpretation. J. Legal Studies 38(2):533–567

Brady PS (1972) The slave trade and sectionalism in South Carolina, 1787–1808. J South Hist 38(4):601–620

Brennan G, Buchanan JM (1985) The reason of rules: constitutional political economy. Cambridge University Press, Cambridge

Browning EK, Browning JM (1994) Public finance and the price system 4th edition. Prentice-Hall, Englewood Cliffs

Buchanan JM (1975) The limits of liberty: between anarchy and Leviathan. University of Chicago Press, Chicago

Buchanan JM (1991a) From private preferences to public philosophy: the development of public choice. In: Buchanan JM (ed) Constitutional economics. Blackwell, Cambridge

Buchanan JM (1991b) Interests and theories in constitutional choice. In: Buchanan JM (ed) The economics and ethics of constitutional order. University of Michigan Press, Ann Arbor

Buchanan JM, Tullock G (1962) The calculus of consent: logical foundations of constitutional democracy. University of Michigan Press, Ann Arbor

Buchanan JM, Vanberg V (1989) A theory of leadership and deference in constitutional construction. Public Choice 61(1):15–27

Curtis GT (1858) History of the origin, formation, and adoption of the constitution of the United States; with notices of its principal framers, vol. 2. Little Brown, Boston

Dorn JA (1991) Madison's constitutional political economy: principles for a liberal order. Constit Polit Econ 2(2):163–186

Du Bois WEB (1896 [1969]) The suppression of the African slave-trade to the United States of America, 1638–1870. Louisiana State University Press, Baton Rouge

Farrand M (1904) Compromises of the constitution. Am Hist Rev 9(3):479–489

Farrand M (ed) (1911a) The records of the federal convention of 1787, vol 1. Yale University Press, New Haven

Farrand M (ed) (1911b) The records of the federal convention of 1787, vol 2. Yale University Press, New Haven

Finkelman P (1987) Slavery and the constitutional convention: making a covenant with death. In: Beeman R, Botein S, Carter EC II (eds) Beyond confederation: origins of the constitution and American national identity. University of North Carolina Press, Chapel Hill, pp 188–225

Finkelman P (2002) The proslavery origins of the electoral college. Cardozo Law Rev 23:1145–1157

Fogel RW, Engerman SL (1974) Time on the cross: the economics of American Negro slavery. Little Brown, Boston

Goldfarb SJ (1994) An inquiry into the politics of the prohibition of the international slave trade. Agric Hist 68(2):20–34
Irwin DA (1996) Three simple principles of trade policy. AEI Press, Washington DC
Jensen M (1964) The making of the constitution. Van Nostrand, New York
Jillson CC (1981) Constitution-making: alignment and realignment in the federal convention of 1787. Am Polit Sci Rev 75(3):598–612
Jillson CC (1988) Constitution making: conflict and consensus in the federal convention of 1787. Agathon Press, New York
Jillson CC, Eubanks CL (1984) The political structure of constitution making: the federal convention of 1787. Am J Polit Sci 28(3):435–458
Kelly AH, Harbison WA (1970) The American constitution: its origins and development, 4th edition. WW Norton and Company, New York
Lerner AP (1936) The symmetry between import and export taxes. Economica 3(11):306–313
Lowenberg AD, Yu BT (1992) Efficient constitution formation and maintenance: the role of "exit". Constit Polit Econ 3(1):51–72
McGuire RA (1988) Constitution making: a rational choice model of the federal convention of 1787. Am J Polit Sci 32(2):483–522
McGuire RA (2003) To form a more perfect union: a new economic interpretation of the United States constitution. Oxford University Press, Oxford
McGuire RA, Ohsfeldt RL (1984) Economic interests and the American constitution: a quantitative rehabilitation of Charles A. Beard. J Econ Hist 44(2):509–519
McGuire RA, Ohsfeldt RL (1986) An economic model of voting behaviour over specific issues at the constitutional convention of 1787. J Econ Hist 46(1):79–111
McGuire RA, Ohsfeldt RL (1997) Constitutional economics and the American founding. Res Law Econ 18(1):43–171
McGuire RA, Van Cott TN (2002) The Confederate constitution, tariffs, and the Laffer relationship. Econ Inq 40(3):428–438
Mcmillin JA (2004) The final victims: foreign slave trade to North America, 1783–1810. University of South Carolina Press, Columbia
Phillips UB (1918 [1966]) American Negro slavery: a survey of the supply, employment and control of Negro labor as determined by the plantation regime. Louisiana State University Press, Baton Rouge
Rawls JA (1971) A theory of justice. Harvard University Press, Cambridge
Robertson DB (2005) The constitution and America's destiny. Cambridge University Press, Cambridge
Shepherd JF, Walton GM (1972) Shipping, maritime trade, and the economic development of Colonial North America. Cambridge University Press, Cambridge
Walton GM, Rockoff HR (2002) History of the American economy. South-Western/Thomson Learning, Mason
Warren C (1928) The making of the constitution. Little Brown, Boston

Chapter 3
A Paradox of Secessionism: The Political Economy of Slave Enforcement and the Union

Phillip W. Magness

Abstract Drawing upon insights from public choice political economy and an examination of historical records, this paper posits an explanation for the causes of secession by the original seven members of the Confederacy in 1860–1861. Secession is examined as a Hirschman exit, intended primarily to shore up and secure the waning federal subsidies and enforcement expenditures that had been afforded to plantation slavery in previous decades. Fears over the impending decline of these subsidies and protections explain the decision to withdraw from the Union, even though slavery itself was, legally, "much more secure in the Union than out of it," to quote Confederate Vice President Alexander H. Stephens. The premises of secession are most evident in southern declarations complaining of the non-enforcement of the Fugitive Slave Act, the instigation of slave insurrections, and the decline of southern political clout. These emphases suggest the perceived threat to slavery was more readily realized in its legal enforcement than in the oft-emphasized territorial question.

3.1 Introduction

"I consider slavery much more secure in the Union than out of it" noted Alexander Stephens in a July 1860 letter to J. Henly Smith (Toombs et al. 1913, p. 487). Consistent with this view, the future Vice President of the Confederacy argued strongly against secession in his home state of Georgia before eventually acquiescing to political tides and casting his lot with the South. Stephens' observation nonetheless reveals a paradox of counterfactuals in the Confederate strategy as the secession movement, undertaken in defense of slavery, actually sparked the events that hastened the institution's demise.

P. W. Magness (✉)
Berry College, Mount Berry, GA, USA
e-mail: pmagness@berry.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_3

Much more could be said—and contested—about the course of the nation's subsequent descent into armed conflict, the political instigations and motives that shaped the course of the Civil War, and the conflicted and sometimes ambiguous role of emancipation as both a Union war aim and military strategy. Yet it also appears readily evident from hindsight that the secessionists miscalculated, and in a way that directly and rapidly undermined their stated motive of preserving slavery such that the institution itself was abolished in barely 4 years time.

The severity of this miscalculation, while generally acknowledged, has proven something of a quandary for historians to explain—or at least in the sense that no clear consensus has emerged to reconcile the mild, gradualist brand of antislavery politics found in Abraham Lincoln's presidential candidacy with the dire predictions of secessionists who pursued separation from the belief that the future of slavery itself was at stake. One common approach credits secession to an increasingly inevitable clash of irreconcilable principles around the national impasse of slavery. Political actors in the "irrepressible conflict" are not always rational agents, and sometimes find their objectives engulfed in a national breakdown of political dialogue or even constitutional institutions that previously functioned to keep secessionism at bay.[1] A second and somewhat more novel explanation actually treats the Republican "threat" to slavery as being correctly diagnosed by the pro-slavery radicals, finding a subtle and certain emancipationist undercurrent in Lincoln's ostensibly moderate "containment" policy of impeding the spread of slavery to the territories even as he disclaimed any interest of interference within the South.[2] By constricting slavery to its existing footholds, so the thesis goes, the institution would experience increasing economic isolation, free labor competition, and an eventual decline of the internal slave trade.

Elements of truth accompany these explanations, though they yield several unsatisfactory answers when encountering particulars of disunion that do not settle easily with their precepts. The "irrepressible conflict" thesis, in its many variants, functions well in describing the angry proclamations of secessionist Fire Eaters and their abolitionist adversaries, yet it ultimately requires their mutual elevation to the political center of each government, and with it an unparalleled suspension of political rationality for frenzied and rash actions with highly uncertain outcomes, sometimes undertaken haphazardly. Political "breakdown" explanations similarly account well for the stalemate of the 1860 election and the absence of a workable compromise around slavery in the vein of prior decades, but only with some susceptibility to a view that reduces secession and the ensuing conflict to a foregone

[1] McPherson (1988, p. 241), for example, presents a picture of secession that tends toward frenzied irrationality, pursued in haste and "carried away by an excess of Robespierrian zeal." A somewhat related thesis emphasizing a destructive turn in the political discourse appears in Varon (2008). An effective and insightful presentation of the "political breakdown" thesis may be found in Holt (1983, 2005).

[2] For a succinct exploration in Civil War causality that examines the role of Republican antislavery pressures in the coming of the war see Levine (2005). An argument in favor of early emancipationist objectives in the Republican electoral and war aims appears in Oakes (2012).

conclusion stripped of the agency of its political instigators. The very notion that secession itself necessarily leads to the Civil War, as experienced, is a causal non-sequitur, even as we may acknowledge that the particulars of secessionism in 1860–1861 prompted specific military actions and responses that precipitated the ensuing conflict and shaped the war's eventual course in ways that were little understood at the time to its primary actors.[3]

Turning to the more novel approaches situated in the Republican "containment" strategy, it is fair to note that the alleged subtle abolitionism attached to Lincoln's election reflects accurately upon a more energetic antislavery presence in the abolitionist wing of the Republican Party. Yet when pressed, such Lincoln-as-antislavery-instigator arguments often require an esoteric reinvention of Lincoln's own gradual and distinctive opposition to the institution in directions of a more proactive design that strain with its particular components and that Lincoln himself would have quite plainly disavowed.

Such discussions merit greater attention than the scope of this study permits, and neither should their insights be neglected from the discussion of Civil War causality on account of the criticisms offered herein. Rather, it is the purpose of this article to suggest an explanation for the aforementioned paradox of pro-slavery secessionism hinted at by Stephens and wrestled with by historians ever since, in this case rooted in political economy. Briefly summarized, the 1860–1861 secession movement may be interpreted as a calculated political exit in the wake of the 1860 election outcome, pursued primarily in recognition that the slaveholding interests had lost their ability to enlist the tools, force, and fiscal resources of the federal government toward the continued maintenance and support of the slave-based plantation economy. By and large, the proponents of this exit strategy did not anticipate an alternative scenario under Lincoln wherein slavery would be immediately imperiled on a legal or constitutional basis, their rhetoric notwithstanding. Rather, they saw an impending loss of institutional controls, expenditures, extracted rents, and regulations that had sustained and subsidized slave-based production as national policy in the decades prior.[4] Isolated from previously enjoyed federal resources and relegated in political voice to a minority position, these secessionists saw no further benefit to continued

[3]It is difficult to understate the sheer uncertainty that shrouded the events of April 1861. Neither the Lincoln administration nor the Confederates accurately anticipated the prolonged conflict that followed Fort Sumter. Gideon Welles, Lincoln's Secretary of the Navy and one of only two cabinet members who remained for the duration of the conflict, candidly conceded in an unpublished 1875 essay that a full "one third of the administration of Mr. Lincoln expired before he had a clear and well-defined policy as to the course to be perceived on important questions affecting the government and the country (Welles 1875, WE 408)."

[4]The political dimensions of a Hirschman (1970) exit may be readily witnessed in the secessionists' actions. Finding an impending diminishment of room for pro-slavery political action within the Union, the original seven Confederate states retreated and attempted to reconstitute themselves as a separate political entity. Ironically, their "exit" action simultaneously deprived the pro-slavery position of its previous near-parity "voice" in the federal government, thereby allowing the same diminishment of political support for the slave status quo that they cited as a reason to secede.

union and accordingly moved in a direction they believed would protect and restore the political entrenchment of slavery that economically sustained its productive output.

3.2 The Political Economy of Slavery

"In every country where the unfortunate law of slavery is established, the magistrate, when he protects the slave, intermeddles in some measure in the management of the private property of the master; and, in a free country, where the master is perhaps either a member of the colony assembly, or an elector of such a member, he dare not do this but with the greatest caution and circumspection (Smith 1978, p. 669–670)." In this brief passage Adam Smith introduced a distinctive feature of the political economy of slavery: its tendency to become politically entrenched. In this sense slavery attains a Stiglerian capture of the labor market, extending itself the protections and financial assistance of the state and erecting political barriers to competitor entry or competitor-induced regulatory oversight.

Though hardly the only wrong of slavery, its fundamental immorality being an entirely sufficient judgment against the institution, political entrenchment does help to explain its persistence.[5] This characteristic is true of slavery even more so than its economic profitability, which is necessarily contingent upon a political order that would sustain its existence and legal sanction. Thus did the celebrated Somerset case note, "The state of slavery is of such a nature, that it is incapable of now being introduced by Courts of Justice upon mere reasoning or inferences from any principles, natural or political; it must take its rise from positive law."[6]

Beyond simply giving the institution legal sanction, state support for slavery is most directly realized in its effecting of the prescribed legal condition. This occurs through the coercive powers of enforcement: the establishment of laws and procedures to return and penalize runaway slaves, of laws that preclude slave uprisings and prohibit material assistance to slaves who would engage in the same, and the supply of a policing agency to administer and execute the laws on both. As Gordon Tullock puts it, "the continuance of the slave system required a very sizeable expenditure on 'security' against the slaves (Tullock 1967, p. 8)." Slaves possess intrinsic human agency even where it is proscribed by law and are kept in slavery only as far as a society enforces it upon them, meaning a slave society must also actively spend resources on coercive suppression to preclude a mass slave exodus, revolt, or both. Furthermore it is a tendency of slaveowners who possess

[5]The classic book *The Slave Power* (Cairnes 1862), for example, presents economically grounded variations of both an inefficiency argument around the slave labor system and a containment argument for its eventual demise. These weaknesses were subjected to and modified by the political ascendency of slave interests.

[6]Somerset v. Stewart. Lofft 1, 98 ER 499 (1772).

political power to diffuse the costs of this enforcement collectively upon society in the process of enlisting the aid of law to their institution (Hummel 1996, pp. 47–61).

The particulars of this "security" or enforcement mechanism are abundantly evident in the American experience. The clauses of the Fugitive Slave Act of 1850 created strict obligations for the U.S. Marshal service to aid in the execution of recapture, while also imposing large criminal fines and penalties upon persons who impeded the return of a slave to his or her owner. The act's notorious 6th section established procedural exemptions and impediments to the execution of due process upon persons seized under the act, effectively subordinating the civil courts themselves to a particularized administrative enforcement. It also financially incentivized the practice of hunting slaves through financial bonuses for successful recapture, and severe fines for persons found to be aiding a runaway.

The expenses of policing and enforcement were similarly shifted onto the general public through the military and federal law enforcement agencies such as the U.S. Marshals. Though it stands out as an atypically severe case, the 1854 recapture of Anthony Burns—Boston's *cause célèbre* of abolitionist resistance—was estimated to have cost the government in excess of $40,000 owing to the deployment of 2000 troops, marshals, and law enforcement officers to escort a single fugitive slave to his owner amidst a riotous mob of opposition. An earlier Massachusetts case, the 1851 rendition of Thomas Sims, came with a claimed $10,000 in extra security (Ericson 2011, p. 91).[7]

These massive expenditures on individual cases were actually touted by some abolitionists as evidence of the immense costs they had imposed upon the enforcers of the law, though a more unsettling realization also lurked behind the price tag. It directly signaled the extreme lengths the federal government was willing to go to return a captive slave, including a show of force that was intended to dissuade future disruptions of fugitive hunting. The collective enforcement of this policing power accordingly functioned as a national subsidy to the slave economy, shifting the burdens of controlling slave escape and resistance onto the public tax role. Without it slaveowners would have had to bear that cost themselves, severely altering the ostensible profitability of the plantation system by internalizing enforcement expenses that were previously shifted and diffused onto the public at large. "The runaway slave," notes Jeffrey Hummel, "was the system's Achilles heel (Hummel 1996, p. 52)."

The Fugitive Slave Act imposed as-of-then unparalleled strains upon federal marshals within the U.S. district court system, charging them with carrying out slave renditions amidst increasing public backlash against the unpopular act (Calhoun 1989, pp. 82–93). Campbell (1970), in his study of the law, records some 332 federal rendition cases, including 68 cases where the federal government paid the

[7]The Burns recapture in particular had a pronounced resonance in the abolitionist movement—in Boston, likely more so than the better known Kansas-Nebraska Act of the same year. See, for example, Thoreau (1854).

entire amount of returning the fugitive.[8] Due to inconsistent reporting, the full dollar value of federal expenses on fugitive enforcement may never be known. A more expansive accounting of fugitive slave enforcement from 1789 to the Civil War might conceivably include federal court and commission hearing costs, a portion of each marshal's time and salary, the expenses of jailing the fugitive, transportation costs, guarding the fugitive including the hiring of additional deputies in the frequent case of public protest, and even diplomatic efforts abroad tied to rendition and compensation demands made upon slaves that escaped beyond the borders of the United States.[9]

In one of the most extensive accountings of government support for slavery to date, Ericson (2011) identifies roughly $51,000,000 in direct federal slavery-related expenditures prior to the Civil War.[10] This figure is undoubtedly the proverbial tip of a much larger iceberg when one considers the unstated costs of other policing and "security" functions that worked to the benefit of the slave system. To this end, the U.S. army as well as state militias functioned as dormant yet easily activated buffers against slave revolts and abolitionist incursions such as the John Brown raid in 1859. Federal arsenals contained weapons that could be deployed to the same end, and federal fortifications across the South housed supplies that could be enlisted as well as the men themselves. Federal marshals had the ability to deputize men to their service or raise a posse to effect renditions, not to mention the assistance of local police and court officers. Even basic federal post office operations contributed to the intercept of "seditious" materials, sent southward by abolitionists.

These and other "security" expenses around slavery, both realized and latent, created additional complications to the antebellum political system by further shifting the "costs" of emancipating slaves onto the general public. This not only entrenched the slave system's operations but imposed barriers to its political divestment. Enforcement over a 4 million person slave population is a costly proposition in manpower alone, to say nothing of extreme cases such the aforementioned

[8]Table 12. Campbell (1970) also documents over $14,000 in federal expenses directly tied to the Burns rendition of 1854 on p. 130.

[9]For example Treasury records denote $18,759 in expenses on fugitive slave renditions between 1850 and the Civil War, yet this figure was inconsistently reported between budget cycles and in some years was omitted entirely. Ericson (2011) utilizes estimates from the aforementioned Sims and Burns cases, and a small number of other high-profile renditions, to extend this number by a conservative $100,000. The actual dollar amount, including hundreds of lesser known cases, is buried under a decade of administrative expenses for the marshal service and federal courts.

[10]Ericson (2011) in his Appendix C figures includes about $9 million in the better-documented expenses of suppressing the international slave trade. This action limited the growth of the institution, though with some irony it also likely contributed to price increases for slaves already in the United States and fostered an expansive internal slave trade. Ericson also assigns $30 million in expenses from the Second Seminole War on account of its close connection to the recovery of fugitive slaves that were harbored within the tribe, as well as additional expenses tied to slave recovery during the Indian removals of the 1830s. Other major federal expenses included about $1 million in recorded army uses of slave labor between 1828 and 1860, debt assumption expenses tied to the annexation of Texas, and the aforementioned fugitive slave renditions.

Burns incident or the military suppression of an active slave revolt.[11] The color line of slavery accordingly became its own cost-control mechanism, particularly as the legal definition around slavery began to shed its somewhat fluid connections to colonial era indentured servitude and as morally instigated manumission practices rose among that most paradoxical creature of the revolutionary era, the antislavery slaveowner.[12] When slavery is legally defined by the slave's skin color, racial attributes also function as an enforcement mechanism to identify suspected runaways. A rise of severe and constricting legal burdens upon manumission accordingly serviced this cost-controlling objective by inhibiting, as (Tullock 1967, p. 8) notes, the "development of concentrations of negroes who were free." Large free black populations in a slave society could both shelter runaways by absorbing them and incubate slave resistance or revolt.

In some respects the peculiar antebellum project of colonization sought a workaround to this enforcement-imposed obstacle to manumission and gradual emancipation. Always a favored policy of antislavery moderates, colonization drew its main support from (1) slaveowners of the Upper South who possessed varying degrees of moral aversion to their inherited institution, i.e. James Madison, Henry Clay, and John Randolph, and (2) antislavery northerners who eschewed the radical, politically disruptive, and extra-legal methods of immediate abolitionism, i.e. Mathew Carey, Benjamin Coates, much of the northern Whig party, and at the time of the Civil War such antislavery moderates as Abraham Lincoln (Magness 2015). However impractical it may now seem, colonization's adherents pursued the tactic of attempting to wean the country from slavery by encouraging gradual manumission and pairing it with subsidized relocation so as not to imperil the existing enforcement mechanisms of the color line. As Clay (1847) himself would note, colonization "obviated one of the greatest objections which was made to gradual emancipation," that being the "continuance of the emancipated slaves among us." Following these parameters, Lincoln's own approach to the slave

[11] A telling and intentionally provocative acknowledgement of the enforcement mechanisms required to sustain the slave system may be found in Henry Highland Garnet's "Address to the Slaves of the United States": "Let your motto be resistance! resistance! RESISTANCE! No oppressed people have ever secured their liberty without resistance. What kind of resistance you had better make, you must decide by the circumstances that surround you, and according to the suggestion of expediency. Brethren, adieu! Trust in the living God. Labor for the peace of the human race, and remember that you are FOUR MILLIONS (Garnet [1843]1990, p. 197)."

[12] Several incidents from early republic Virginia signify the shifting nature of the color line as a mechanism of slavery enforcement. The gradual liberalization of the state's manumission procedures in the post-revolutionary era, perhaps best signified by Robert Carter III's famous 1791 "Deed of Gift," gave way to increasing legal restrictions around the process in the wake of Nat Turner's rebellion coupled with increasingly stringent stipulations to prompt the relocation of free blacks from the state by the eve of the Civil War. The early "freedom suit" of Hudgins v. Wright (1806) similarly raised the blurred lines of slavery and indenture in the case an enslaved family of mixed racial and part-Indian lineage. Though the case initially prevailed in state district court, a partial reversal at the appellate level stratified slavery along racial lines in differentiating black and Native American ancestries.

issue articulated both in his pre-presidential career and pursued in office was consistently linked to the dual proposition of gradual compensated emancipation and colonization abroad on a voluntary but subsidized basis.

Slavery's institutional situation at the outset of the Lincoln presidency could accordingly be characterized as the product of many decades of political entrenchment and subsidy at the national level—features that also permitted its economic viability. As the instigator of the secession crisis, the election of Lincoln might also be construed as a perceived or real threat to the political status quo. The means by which this event specifically precipitated the secession of at least the lower South, and on what decision basis they attempted an exit may be found in the implications of the election outcome not for slavery itself but for its previously enjoyed benefits vis-à-vis a perceived gain from starting out alone.

3.3 What Lincoln's Election Meant for Slavery

On the eve of the 1860 Republican Convention, New York Tribune editor Horace Greeley confided his prognostication on the approaching electoral contest to an associate. "I want to succeed this time, yet I know the country is not anti-Slavery. It will only swallow a little anti-slavery in a great deal of sweetening. An anti-slavery man per se cannot be elected; but a tariff, river and harbor, Pacific railroad, free homestead man may succeed although he is anti-slavery."[13] While he may have had other prospective candidates in mind at the time, Greeley's observation fit well with the man who emerged from the convention as the Republican's standard bearer.

Abraham Lincoln was indeed an anti-slavery man, though of a distinctively Whiggish and gradualist flavor. He came from the mould of Henry Clay and espoused Clay's own formula of "compensated emancipation and colonization" as a means of parting the country from its peculiar institution. He adhered to the Republican line of excluding slavery from the territories, though he also pledged non-interference with the institution where it already existed save to incentivize its gradual dissolution through the aforementioned means. The lawless tendencies of a more radical abolitionism clashed with his conceptualization of civil discourse in a political society. Notably, Lincoln also embodied the logrolled assemblage of Whig socio-economic policy outlined by Greeley's letter (Luthin 1944).

Considered in full, Lincoln's platform directly mirrored and continued the political legacy of the antislavery slaveholder Henry Clay—a point that he repeatedly and publicly broadcasted dozens of times between the famous Lincoln-Douglas debates of 1858 and his election as president. While the old "Whig formula" of compensated emancipation and colonization had occasionally fostered disunionist bluster from the political periphery of the South in prior instances, it was hardly the John Brown-infused rhetoric of immediatist abolitionism and rested not far from the

[13] Horace Greeley to Margaret Allen, January 6, 1860.

political center of the late antebellum. It also operated within the constraints created by the aforementioned political entrenchment of slavery. Specifically, Lincoln's formula pledged a plausible extraction from slavery—even if impractical by modern standards—that did not disrupt its existing enforcement and only slowly pressed against its continuation.

Before turning to the question of how an outwardly moderate candidate, not far removed from the prior generation's conciliatory standard-bearer, could provoke an aggressive secessionist exit, it is worth mentioning an under-acknowledged concession to slavery that Lincoln did make at the outset of his presidency. In late February 1861, Congress hastily adopted a proposed stopgap measure called the Corwin Amendment to affirm explicit constitutional sanction upon the "domestic institution" of slavery in states where it already existed. This constitutional pledge of non-interference was intended to stave off secession and lure back its early adherents in the Deep South. The amendment's exact legislative origins are concealed by the heavy politicking that produced it, yet Lincoln's personal support for the measure is surprisingly easy to establish given the scant scholarly attention paid to the proposal (Crofts 2014).[14] After a brief flash of congressional support at the close of the "secession winter" Congress and a public nod in Lincoln's inaugural address, the Corwin Amendment was quickly overtaken by the events of the war and failed to attain ratification. As an intended bone for the South it accomplished little and proved insufficient to stem the secessionist tide, though its larger constitutional implications might be the manner in which it strafed the old radical abolitionist debate about the Constitution's disputed pro or anti-slavery character.[15]

What more pressing issues than the offer of explicit constitutional recognition for slavery motivated southern secessionism then? The answer derives from the aforementioned considerations of political economy, and the evidence is abundant in a number of well known yet lightly-scrutinized causal declarations from the initial seceding states. Stated simply, whereas Lincoln's election posed no plausible direct threat of effecting abolition, rhetoric to the contrary notwithstanding, the larger Republican victories of 1860 did portend a shift in political power that weakened the political entrenchment of slavery. Prior to the election, the captured tools of government permitted the slaveowning interests of the South to offload the costs and mechanisms of slavery's enforcement upon the whole of the nation and its treasury. Despite his constitutional concession to slavery's existence, Lincoln's election portended a weakened commitment to its public subsidization. Secession may

[14] See also, Magness (2014). Lincoln's orchestration of the Corwin Amendment is further affirmed in a letter Duff Green to the editor of the New York World on September 26, 1865, republished in the *National Intelligencer*, October 5, 1865. According to Green, Lincoln revealed this role to him in a conversation aboard the USS Malvern during his April 1865 visit to the surrendered city of Richmond.

[15] On this point, abolitionist Lysander Spooner noted that the Corwin Amendment "was an admission that, as [the Constitution] then stood, it did not recognize [slavery]. And if it did not, blacks were citizens." Spooner to Henry Summer, November 30, 1866, Special Collections, Bryn Mawr College.

accordingly be viewed as a response to the impending loss of a political-regulatory capture, the "benefits" to slavery provided by that capture deriving not from its strict legal or constitutional status but from an ability to muster public resources to its aide and to the alleviation of the enforcement costs of its perpetuation.

3.4 The Secessionist Impulse

Secessionist perception of their forthcoming political isolation became immediately apparent after the 1860 election. "As to this body, where do we stand?" thundered Senator Louis T. Wigfall in a challenge before the chamber on December 20, 1860. The pro-slavery 'Fire Eater' from Texas proceeded to lay out the political impotence of the slave states under the incoming administration by running through the numbers of the Senate, which until that point had been something of a stopgap against the numerical population advantages the North enjoyed in the House:

> Why, sir, there are now eighteen non-slaveholding States. In a few weeks we shall have the nineteenth, for Kansas will be brought in. Then arithmetic which settles our position is simple and easy. Thirty-eight northern Senators you will have upon this floor . . . There are four of the northern Senators upon whom we can rely, whom we know to be friends, whom we have trusted in our days of trial heretofore, and in whom, as Constitution-loving men, we will trust. Then we stand thirty-four to thirty-four, and your Black Republican Vice President to give the casting vote. Mr. Lincoln can make his own nominations with perfect security that they will be confirmed by this body, even if every slaveholding State should remain in the Union, which, thank God, they will not do. You have elected your President, and you can inaugurate him; and we will have neither lot nor parcel in this matter.[16]

For all their fuming disunionist threats, Wigfall's remarks conveyed an underlying rational calculation about the political landscape before him. Slavery only enjoyed its privileged position from the political support it mustered, and now two previously reliable buttresses for that support—the presence of a friendly president to appoint favorable administrators and the numerical parity of the Senate to sustain a favorable legal status for enforcement—were no longer available. It mattered not that Lincoln pledged non-interference with slavery where it existed, that he backed a constitutional amendment to affirm as much, or that he even acquiesced to the execution and enforcement of the Fugitive Slave Act. Slavery had lost its political entrenchment such that an effective political defense of the previously-enjoyed subsidy mechanisms could no longer be mounted, much less an extension of their proactive enforcement against state-level political institutions of the North that increasingly turned to open defiance of the federal statute in the late 1850s.

The motives of secessionism must be considered in its waves, as the Upper South followed this course upon the outbreak of the Civil War and moved in direct reaction to the events after Fort Sumter (Crofts 2014). Their motives accordingly diverged in complex particulars from the lower South and should be evaluated in the context of

[16]Congressional Globe, 36th Congress, 2nd Session, p. 75.

the unfolding military events that prompted them. The seven states of the original secessionist wave justified their actions on the cause of opposing Lincoln's election though, and published several statements explaining their chosen course. These "secession declarations" are accessible and well known, although the study of their contents has, at times, shown a tendency to succumb to simplified analysis around the readily apparent, though superficially considered, observation that slavery lies at the core of each.[17]

South Carolina, the first mover of the secession crisis, actually published two justifications of its course in a written "declaration of causes" and a separate but parallel "address" laying those causes before the other slaveholding states. Georgia, Mississippi, and Texas each followed suit with declarations of their own. The Florida secession convention also commissioned a draft declaration to serve a similar purpose, though its quorum departed and the meeting adjourned without officially adopting any statement. An early draft of the document nonetheless survived and was recently rediscovered in state archival records.[18]

Excepting the South Carolina "Address," which presents a historical and legal-constitutional argument for secession over reasons still linked plainly to slavery, the other five "Declarations" are distinctive not merely for the primacy of slavery in their articulated causes, but rather in their recurring identification of several particular threats to the slave system's political entrenchment. Almost uniformly, they turn to:

1. Non-enforcement of the Fugitive Slave Act, itself a matter of direct federal sustenance for slavery,
2. The fomentation of slave resistance and rebellion, thereby invoking and questioning the government's commitment to a guarantee of military suppression,
3. The exclusion of slavery from the territories, the last of these being a numerical tool to break the remaining balance of slave state votes in the Senate.

These specific grievances are direct and recurring. Mississippi complained of the negation of "the Fugitive Slave Law in almost every free State in the Union." Georgia identified the constitutional and legal guarantees of the return of fugitives as "our main inducements for confederating with the Northern States. Without them it is historically true that we would have rejected the Constitution." Other clauses charged the federal government and the northern states with failing to police against the instigation of slave insurrections, or to prosecute those who would aid them.

[17]For example, Loewen and Sebesta (2011) is an edited compendium of these documents with interpretive notes. The thrust of their argument, however, is premised upon demonstrating the easily sustained point that slavery, as opposed to other "neo-confederate" revisions, was in fact the primary causal instigator of secession. It is nominally correct on this point, but wholly superficial in the sense that it sacrifices the opportunity for a deeper analysis of the particulars of slavery to a retreading of a debate waged largely in amateur and non-scholarly settings.

[18]The unpublished Florida document was rediscovered in 2012 by historian Dwight Pitcaithley in the papers of Gov. Madison Starke Perry. It appears in an unedited draft in the State Archives of Florida, Series 577, Carton 1, Folder 6.

The draft Florida proclamation cited the refused surrender of "fugitives from justice charged with treason and murder" from the John Brown plot. Georgia pointed to "the murderers and incendiaries who escaped public justice by flight have found fraternal protection among our Northern confederates." Texas charged the government with failing to intercept "seditious pamphlets and papers among us to stir up servile insurrection and bring blood and carnage to our firesides." Taken in cumulative, the secession documents of the original seven Confederate states are not merely a proclamation of their cause's attachment to slavery. They are a litany of specific accusations in which the federal and northern state governments are charged with neglecting, failing, or intentionally abrogating prior commitments to politically sustain and subsidize the policing and enforcement mechanisms that sustained the entire slave system.

By point of comparison:

- All five documents identify the non-enforcement of the Fugitive Slave Act and accompanying constitutional clauses as a core grievance,
- South Carolina and Texas lambast the northern states that have "nullified" fugitive slave enforcement through various circumventing acts of their courts and legislatures, while Georgia cites a specific instance of the same from Wisconsin,
- All five documents attribute the fomenting of servile insurrection to the north, and the harboring of their instigators without legal recourse,
- South Carolina, Mississippi, Georgia, and the Florida draft specifically allude to the John Brown raid,
- All five identify the settling of the territorial question against slavery as a threat to the southern political balance in Congress,
- All five express concerns that the incoming president and Congress will fail to sustain an enforcement policy for fugitive rendition or adequate guarantees against insurrection at the federal level,
- All five reference the reduction of the slave states to a political minority as cause for separation.

Though the heated rhetoric of the documents intimated a foreboding affinity between Lincoln and the abolitionist wing of his party, the documents also evince a subtler recognition that he did not intend to move directly against slavery. Thus the Florida draft concedes that "It is denied that it is the purpose of the party soon to enter into the possession of the powers of the Federal Government to abolish slavery by any direct legislative act. This has never been charged by any one." Rather it was the belief that Lincoln's election portended an accomplishment of this result by long attrition; that it meant executive inaction would supplant federal assistance and enforcement of the laws that sustained slavery or, as Texas' declaration indicated, that Lincoln's election signaled an "approval of these long continued wrongs" against slavery through what it characterized as northern intransigence and legislative subversion at the state level.

Notably, the thrust of the southern grievances over slavery, though aggressively asserted, tend to fall well outside of the scope of the constitutional status question that the Corwin amendment intended to address. They do however reflect a closer

connection to the Crittenden compromise resolutions and those of the Washington Peace Conference mirroring Crittenden—two failed attempts to devise a workable legislative arrangement that would stave off secession prior to Lincoln's inauguration. In addition to measures designed to resolve the territorial balance, Crittenden's package contained a proposed constitutional amendment to strengthen the Fugitive Slave Clause and three separate resolutions to bolster enforcement of the Fugitive Slave Act.[19] That such measures were entertained further confirms that federal enforcement mechanisms, and the political clout to sustain them, were at the center of the secessionist impulse.

3.5 Creating a Slave-Based Security State

The provisions of the Confederate Constitution of 1861 provide another direct window into the secessionist motives of the preceding winter, and with it suggest a series of steps intended to restore and strengthen the advantages the slave system derived from its previous position in the federal government. This document, adopted after the initial secession of the Deep South but before the war-induced departure of the Upper South, specifically replicated the U.S. Constitution of 1787, though with specific modifications pertaining mostly to slavery. Thus the Confederates declared a constitutional "right in property" to slaves including transit and sojourn within its borders, a strengthened Fugitive Slave Clause, and a protection of slavery in all acquired and admitted territories.

By design and intended implementation, the document presented the structures of a slave-based security state. Its offer to existing and prospective members of the Confederacy was the reestablishment of a clear federal absorption of the burdens of enforcing the slave system, and the perpetuated political entrenchment of these features at the constitutional level. The new constitution sought to reverse the weakened enforcement structures, both real and anticipated, of the very same institutions in consequence of the 1860 election, and present them as the collective benefit of a reconstituted political arrangement. Notably, this arrangement also came into existence upon a *de facto* concession of prior legal claims upon the status of the United States territories. While the foregoing of the territories did not preclude Confederate designs for expansion after separation from the United States, it does suggest that the primary function of the antebellum territorial question—often placed at the center of secessionist causality—had more to do with the numerical balance in the Senate than the commonly encountered "containment" thesis of Republican free-soil strategy. As much as "Bleeding Kansas" foreshadowed the conflict of the following decade, the secessionist acts of the Deep South effectively signaled their own retreat from their hotly contested claims to Kansas or the other territories in exchange for internally strengthened government protections to slavery

[19] Amendments Proposed by Senator John J. Crittenden, December 18, 1860.

where it already existed.[20] To the slave-owning instigators of secession, the benefits provided by collective absorption of the enforcement costs under a new national constitution and government evidently outweighed the projected gains of expanding the same institution into the now-relinquished territories of the United States, at least in the short term.

Whether the Confederacy could have sustained a re-entrenchment of slaveholding interests and effectively held the enforcement costs of the institution at a national level is a separate question. While one secessionist predicted an "impassible wall between the North & the South," the ability to control a new Confederate border, fund the policing mechanisms that would both inhibit runaway slaves, and defend against abolitionist incursions may well have become the Achilles heel that Hummel (1996) identifies.[21] Lincoln suggested his awareness of that much as late as his Second Annual Message of December 1, 1862, holding forth the Fugitive Slave Clause as a final enticement for the southern states to return:

> The fact of separation, if it comes, gives up on the part of the seceding section the fugitive-slave clause, along with all other constitutional obligations upon the section seceded from, while I should expect no treaty stipulation would ever be made to take its place.[22]

The Confederate exit to bolster slavery's support in a new structure of government had the direct effect of foregoing the remnants of support in the old one. As Hummel (2012) notes (p. 365) in a separate economic analysis of slavery, "The individual runaway, in the final analysis, helped both to provoke secession and to ensure that secession was unable to shield slavery in the end."

Was slavery in fact safer in the union than out of it as Alexander Stephens had maintained? Strictly as a matter of its legal grounding connected to an established constitutional system, probably so. The combination of a constitutional order that generally recognized slavery and the offer of a compromise amendment to strengthen and clarify its standing in positive law would have likely sustained the institution for some time to come absent the disrupting events of 1861–1865. The South's abandonment of this claim, combined with the course of militarily-justified decisions during the war, undoubtedly hastened and permitted a federal emancipation policy.

The political economy of slavery's future was far more ambiguous. Slaveowner interests in the winter of 1860–1861 confronted a sudden and rapid decline of their own political clout, which had thus far provided a necessary institutional support for their economic viability by spreading the costs of its sustenance onto the public at large. While legally established and even supported in recent judicial opinions, those institutions were already undercut by growing external pressures, by organized state-level non-enforcement, and even to some degree their own impracticality and

[20] May (2013) makes a strong argument that the late antebellum territorial dispute turned upon southern expansionist designs in the Caribbean, rather than its heavily emphasized interests in the domestic territories of the west.

[21] See pp. 52–56 and 133.

[22] Abraham Lincoln, "Second Annual Message to Congress," December 1, 1862.

unwieldy cost requirements before the election ever came. Barring any prolonged war, or even permitting its expected course to produce a rapid victory in southern favor, the political calculus of secession becomes much more favorable to exit, so long as the nascent secessionists believed they could restore and strengthen the state subsidies and enforcement mechanisms that slavery enjoyed in prior decades. Here the choice in 1860 was not between slavery and abolition, but between slavery increasingly stripped of external support and funding at the national level and slavery reconstituted under a separate national government pledging its explicit subsidy and full energy. With the security of the former diminishing, the latter attained greater political appeal and from the slave-owner's perspective quite rationally so.

The attempted political exit that followed therefore unfolded not along Stephens' terms of the institution's legal status, but in defense of its prior capture of the mechanisms of government to disperse the burdens of its own enforcement onto the public and public treasury. It is the same capture decried in republican governments by Adam Smith, wherein "[t]he persons who make all the laws in that country are persons who have slaves themselves." Following their interests, such persons "will never make any laws mitigating their usage; whatever laws are made with regard to slaves are intended to strengthen the authority of the masters and reduce the slaves to a more absolute subjection (Smith 1978, p. 181)." In the Confederate example, it might be duly added that in the face of diminishing stakes in the existing government, the slaveowner will further move to strengthen the entrenchment and the extracted gains of his institution by separation from that which no longer acts to its political advantage.

References

Cairnes JE (1862) The slave power. Carleton, New York
Calhoun F (1989) The lawmen: United States marshals and their deputies, 1789–1989. Smithsonian University Press, Washington, DC
Campbell SW (1970) The slave catchers: enforcement of the Fugitive Slave Law, 1850–1860. UNC Press, Chapel Hill
Clay H (1847) Speech of the hon. Henry Clay, at the mass meeting in Lexington, Ky., on Saturday, November 13, 1847. Crissey & Markley Printers, Philadelphia
Crofts DW (2014) Reluctant confederates: Upper South unionists in the secession crisis. UNC Press, Chapel Hill
Ericson DF (2011) Slavery in the American republic: developing the federal government, 1791–1861. University Press of Kansas, Lawrence
Garnet HH ([1843]1990) An address to the slaves of the United States of America. In: Ravitch D (ed) The American reader: words that moved a nation, HarperCollins, New York, pp 192–197
Hirschman AO (1970) Exit, voice, and loyalty: responses to decline in firms, organizations, and states. Harvard University Press, Cambridge
Holt MF (1983) The political crisis of the 1850s. Wiley, New York
Holt MF (2005) The fate of their country: politicians, slavery extension, and the coming of the civil war. Hill and Wang, New York

Hummel JR (1996) Emancipating slaves, enslaving free men: a history of the American Civil War. Open Court, Chicago
Hummel JR (2012) Deadweight loss and the American Civil War: the political economy of slavery, secession, and emancipation
Levine B (2005) Half slave and half free: the roots of civil war. Hill and Wang, New York
Loewen JW, Sebesta EH (2011) The confederate and neo-confederate reader: the "great truth" about the "lost cause". University Press of Mississippi, Jackson
Luthin RH (1944) Abraham Lincoln and the tariff. Am Hist Rev 49(4):609–629
Magness PW (2014) Abraham Lincoln and the Corwin amendment. http://historynewsnetworkorg/blog/153391
Magness PW (2015) The American system and the political economy of black colonization. J Hist Econ Thought 37(2):187–202
May RE (2013) Slavery, race, and conquest in the tropics: Lincoln, Douglas, and the future of Latin America. Cambridge University Press, Cambridge
McPherson JM (1988) Battle cry of freedom: the American Civil War. Oxford University Press, Oxford
Oakes J (2012) Freedom national: the destruction of slavery in the United States, 1861–1865. WW Norton & Company, New York
Smith A (1978) Lectures on jurisprudence. Liberty Fund, Indianapolis
Thoreau HD (1854) Slavery in Massachusetts. New York Tribune November 12
Toombs RA, Stephens AH, Cobb H (1913) The correspondence of Robert Toombs, Alexander H. Stephens, and Howell Cobb. American Historical Association, Washington DC
Tullock G (1967) The economics of slavery. Left Right 3(2):5–16
Varon ER (2008) Disunion!: the coming of the American Civil War, 1789–1859. University of North Carolina Press, Chapel Hill
Welles G (1875) Gideon Welles papers, 1846–1902. The Huntington Library, San Marino

Chapter 4
Why Is There a Ratchet Effect? Evidence from Civil War Income Taxes

David Mitchell

Abstract The ratchet effect in public finance refers to the historical phenomena that the size of government increases during a crisis but does not return to its previous level when the crisis ends. The traditional explanation is that voters change their views on the appropriate size of government during the crisis. But change in tastėi is an explanation of last resort: it should not be accepted without examining alternatives. This paper looks Civil War taxes as an illuminating case of the ratchet effect. Both the observed political process and the resulting mix of taxes suggest that interest groups, not voters, led to the ratchet effect in this case. During the Civil War both tariffs and income taxes increased, but only the higher tariff stayed. This paper uses an analytical narrative to show that this was because the new interest groups only wanted the higher tariff and not the income tax.

4.1 Introduction

The ratchet effect is an empirical phenomenon related to the growth of government. It was discovered by Peacock and Wiseman (1961), who found that most government growth was associated with periods of crisis such as wars. They noticed that government expenditures as a percentage of GDP have an upward-sloping trend line, but also that growth of government increases during a crisis and does not return to its former level after the crisis passes. Higgs (1987, p. 50) found a similar result in the United States. Rasler and Thompson (1985) found that the ratchet effect holds internationally. To explain the ratchet effect, Peacock and Wiseman postulated that voters allow governments to raise taxes in order to pay for necessary crisis-related expenditures. When the crisis ends, voters' views of appropriate tax levels have changed. Voters then accept a larger government. Thus, their explanation relies on a change in voter ideas or preferences.

D. Mitchell (✉)
University of Central Arkansas, Conway, AR, USA
e-mail: dmitchell@uca.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_4

As they put it, outside of crisis periods:

> Notions about taxation are likely to be more influential than ideas about desirable increases in expenditure in deciding the size and rate of growth of the public sector. There may thus be a persistent divergence between the ideas about desirable public spending and ideas about the limits of taxation. This divergence may be narrowed by large-scale social disturbances, such as major wars. Such disturbances may create a displacement effect, shifting public revenues and expenditures to new levels (Peacock and Wiseman 1961, p. xxiv).

Taxation during the U.S. Civil War provides an interesting counter-example to Peacock and Wiseman's account. During the Civil War, government spending increased dramatically and exhibited a ratchet effect afterwards. After the war, however, the income tax was completely eliminated while the tariff was merely reduced. The traditional view of the ratchet effect would suggest that voters in the 1860s became accustomed to higher levels of taxation on both income and imported goods. Instead, I find that interest groups organized legislation that continued to increase the tariff rate throughout the Gilded Age while repealing the income tax. I make my case through an analytical narrative of the process that implemented, then removed, the Civil War-era income taxes.

4.2 The Ratchet Effect in the US Civil War

During a crisis, several things happen. The perceived benefits of various government expenditures change. What had been an effective policy equilibrium is no longer. The perception that a different policy equilibrium is more suited to dealing with the crisis provides an incentive to change the prevailing structure or combination of existing interest groups and the rules that protect them. Groups poised by chance to take advantage of the situation do so. They strengthen their position during the crisis and attempt to get key coveted positions after the crisis. Idiosyncrasies particular to the crisis encourage voters to insist on specific structural changes. But once the previous structure is broken, many of the rules that supported the previous structurally induced equilibrium are no longer in effect. Thus, the relative cost of moving to any new policy equilibrium decreases dramatically, particularly the cost of being at a larger level of government.

4.2.1 The Crisis and Immediate Response

Prior to the Civil War, internal federal taxes barely existed. Government played a small role in the economy. Tariffs and the sale of public lands provided most federal revenue. From 1817 to 1857, the federal government did not use the excise, stamp, income or property taxes (Smith 1914). The national debt was small. During most years the federal government ran a surplus. Even during the Mexican-American

War, the U.S. had no need for internal taxation. Occasional shortfalls in revenue were met by issuing Treasury notes.

When South Carolina attacked Fort Sumter on April 12, 1861, however, the need for revenue suddenly increased. Shortly afterward, Lincoln called for 75,000 new militiamen and summoned a special session of the Congress to meet on July 4. In order to fight the war, the Union needed to raise revenue and to borrow large sums of money. The size of the federal government increased dramatically during the war. As Holcombe (1999, p. 13) notes, "the magnitude of post-Civil War expenditures was at least double pre-war outlays." Government receipts as a percentage of GDP stayed well above its antebellum rate even as the GDP was increasing (Johnston and Williamson 2003). But before an income tax could come into being, a variety of groups had to organize and find a new tax policy equilibrium. The actual process that brought about the increase in taxation provides important evidence that interest groups, not voters, were critical to the ratchet effect.

4.2.2 The Process of Forming a Coalition

Because the Southern states were no longer represented in Congress during the Civil War, the Republican Party had a large majority. However, this did not give Republican leadership carte blanche. The loyalty of the citizenry was suspect in several regions of the Union.[1] Implicitly understanding Olson's (1965) theory of interest groups, the leadership retained power and preserved the Union by forming a winning coalition that could keep Border States from seceding, accommodate Copperheads-those Democrats who sympathized with the South, and raise money to fight the war.[2]

The Union leadership needed to keep three disparate groups within their coalition: the yeoman farmers who filled the Midwestern and Border states and provided many of the Union soldiers, the finance capitalists who lent the Union money, and the manufacturers who made up the strength of the Republican Party.[3] Northern Democrats deliberately exploited the tax fairness question, turning taxation into what McPherson (1988, p. 442) calls the second most divisive issue in the North during the war. Many felt that it was a poor man's fight and a rich man's war. Working class people and farmers noted that tariffs raised the prices and profits on manufactured goods.

[1] Tabarrok and Spector (1999) analyze the 1860 election and Lincoln's popularity.

[2] During his tenure Lincoln carefully allowed certain issues to appear to be the work of Congress alone. These included increased tariffs and the national banking system. These issues were sectionally divisive, and Lincoln did not want them to break up his coalition. Copperheads were conciliatory Democrats.

[3] A fourth important group was the Radical Republicans, which overlapped the other three groups but the loyalty of the Radical Republicans was never questioned. Radicals were inelastic demanders of Republican leadership.

Border States were also sensitive to the tax issue. Lincoln was very keen to form a coalition with moderates and War Democrats, in part because he was worried about the possible secession of the Border States. To keep these states in the Union he needed at least to keep up the appearance that the burden of the war was being shared equitably. Equity was especially relevant because by the time the income tax finally passed on August 5, 1861, the North had already lost at Bull Run. The popularity of the war was temporarily waning (McPherson 1988, p. 348). This was exacerbated by the feeling that war might not be fought or financed in a manner fair to all.

Eventually most of the cost of the war would be paid with tariff revenue, but tariff revenue was coming in slowly in the middle of 1861.[4] The loss of Southern tariff revenues and an initial disruption of commerce caused tariff revenues to fall dramatically. Tariff revenue fell from $9,772,574 during the previous 3 months to $5,515,552 during the next 3 months. Some Republicans advocated abandoning the gold standard immediately, but this met fierce opposition from hard-money interests. Congress discussed issuing treasury notes. This also met with opposition, and the government decided to issue bonds.

Bonds were difficult to sell, not only because the country was undergoing a massive civil war, but because many Southerners had already dumped their U.S. bonds in England and other European countries.[5] Foreigners were similarly flooding the market with U.S. securities including national, state, and corporate debt (Bensel 1991, p. 249). As a result, demand for new U.S. bonds was low (Dewey 1903). Trying to borrow $150 million, Secretary of the Treasury Salmon Chase went to New York City to meet personally with a group of bankers who might be persuaded to purchase or underwrite federal bonds (Bolles 1886, Vol. II, p. 21).[6]

Borrowing alone (without new taxes) could not raise sufficient funds, in part because financiers had to be convinced that future tax revenues would be large enough to ensure repayment of bond debt. Lenders were already wary. Thaddeus

[4]Secretary of the Treasury Salmon P. Chase proposed that uninhibited lands in the northwest and west could be sold, but such sales would not bring revenue to the Union for quite some time.

[5]Banks were also having difficulty selling bonds in part because of the Trent Affair. The Trent Affair occurred in Nov. 8, 1861 when the U.S.S. San Jacinto boarded the British mail carrier Trent and took two Confederate emissaries by force. Britain was understandably upset by the incident. The U.S. was forced to apologize and return the prisoners. Bondholders were afraid that if war with Great Britain broke out, the chances of the U.S. ever repaying debt would be dramatically reduced.

[6]There are several accounts of Chase's difficulty in selling the bonds, including Sharkey (1959). The Independent Treasury Act of August 1846, which stated that the Treasury could pay out and accept payment only in coin, proved a major impediment. Chase was able to get this law amended in August of 1861, but the delay made raising revenue quite difficult. Chase met with representatives of thirty-nine New York banks as well as representatives of Boston and Philadelphia banks. These bankers would later be an important interest group within the Republican Party. Chase focused on eastern banks because western banks lacked the resources to purchase or underwrite bonds in any great quantity (Dewey 1903).

Stevens (R-PA, 9th District), chair of the House Ways and Means Committee, declared in a debate on July 24, 1861:

> The capitalists must be assured that we have laid taxes which we can enforce, and which we must pledge to them in payment of the loans, or we shall get no money. (Stevens 1997, p. 216)

With tariff revenue already apparently maxed out, attention turned to internal taxes. At the beginning of 1861, the government paid little attention to the income tax; taxes on property and on manufacturing appeared more promising. Federal property taxes could be piggybacked onto existing administrative state and local systems, providing an easy method for raising revenue.[7] However, most members of Congress considered a property tax to be a "direct tax." Article 1, Section 8 of the Constitution required the federal government to allocate a direct tax among the states based on population, not wealth or property values.[8] Representatives from poorer states complained that their constituents would face a higher tax rate than constituents in richer states. Furthermore, there was concern that intangible property would not be taxed. Wealthy finance capitalists would pay little if any tax, while farmers who held all their wealth in real property would pay a high tax.

Representative Schuyler Colfax (R–IN) stated:

> I cannot go home and tell my constituents that I voted for a bill that would allow a man, a millionaire, who has put his entire property into stock, to be exempt from taxation, while a farmer who lives by his side must pay a tax.[9]

Colfax was speaking directly of the difficulty in creating a winning coalition, or win-set in spatial-analytic terms. The Republican Congress would need to carefully construct a set of policies in order to keep all the disparate groups involved. Taxes needed to have some appearance of fairness, especially with respect to vertical equity. Colfax could tell his constituents that the crisis called for higher taxation and expenditures, but not that they should bear an unfair share of the burden. Property taxes were defeated, and the leadership looked for new taxes.

4.2.3 The Civil War Income Tax of 1861

The multifaceted tax bill that eventually passed through both houses in August 1861 was a compromise bill that, among other provisions, taxed income above $800 per year at 2%. In order to encourage the purchase of government bonds, the income tax included the first loophole: the tax rate was reduced on income from government

[7] States collected direct taxes for the U.S. government in 1813, 1815, and 1816 (Smith 1914, pp. 24–25).

[8] *Congressional Globe*, 1st Sess., 37th Cong. 1861, p. 252.

[9] *Congressional Globe*, 1st Sess., 37th Cong. 1861, p. 306. Colfax would become Speaker of the House in 1863.

bonds. Congress had some notion of how an income tax would work because some states and municipalities in the U.S. used an income tax—including all the New England states, Boston, and Savannah (Kinsman 1903). In addition, both Congress and bondholders were aware of Britain's success with its income tax.[10]

Despite the fact that this was the first U.S. income tax, other parts of the tax bill—particularly changes in tariff rates—received much more attention.[11] The income tax was expected to bring in a mere $5 million (Smith 1914). The changes in tariffs were expected to provide more revenue and to provide it more quickly.[12] Moreover, Congressmen from the manufacturing districts were quite eager to raise protective tariffs to levels that had been impossible while the Southerners—who cited high tariffs among the reasons for their secession—remained in Congress.

House leadership believed, however, that a low income tax rate was ideal. The tax only needed to generate enough revenue to convince finance capitalists that bonds would be repaid.[13] The mere existence of a tax on the wealthy (even at a low rate) kept the masses happy, including voters in Border States, by providing some semblance of vertical equity (Stanley 1993). Manufacturing interests were happy with the low rate because it did not produce enough revenue to reduce support for tariffs. Nor did it outrage wealthy supporters. The income tax did not seriously impede the groups that put the Republicans in power.

Interestingly, the progressive nature of the income tax bill was not hotly debated. It was part of the issue of vertical equity. More importantly, the progressive tax was thought to bring in more money. That had been the case in Prussia's graduated income tax (Witte 1985; Brownlee 2003).

4.2.4 The Civil War Income Tax of 1862

Raising revenue in 1862 was contentious in Congress for several reasons. Civil War fighting had been much tougher than expected. Casualties at Shiloh alone exceeded

[10] Great Britain enacted the first income tax in 1798 under Pitt the Younger and abolished it in 1816. Sir Robert Peel instituted a second temporary income tax in 1842. However, due to the Crimean War it was not abolished as planned. The Confederacy also used the income tax both at the national and state level. For more on the history of income taxes, see Kinsman (1903) and Seligman (1895).

[11] All writers on the subject note how little attention the income tax received at the time. Tariff schedules and property taxes account for much more of the debates for two reasons. First, tariff revenues would come in sooner. Second, representatives of organized manufacturing groups tended to hold leadership positions with agenda control. This is particularly true of the iron men.

[12] The rate on coffee received the most attention. Chase suggested that coffee and tea be removed from the free list and that the low tariff on sugar be raised.

[13] The finance capitalists were an organized group within the Republican Party. There are a variety of letters to both Houses from organized groups of finance capitalists. These can be accessed electronically at The Journal of Senate of the United States of America and *The Congressional Globe*. The interested reader can also look at Sharkey (1959, p. 250), Bensel (1991), McPherson (1988, p. 599), and Unger (1959).

the number of killed or wounded in all previous U.S. wars combined.[14] After Shiloh, Union leadership realized that the Civil War would not be short; many more men and much more money would be needed. At the same time, citizens were losing confidence in the Union's ability to win the war. When the fighting first began, Union recruiting offices had more volunteers than they sought, but by 1862 recruits were tough to find.

The loss of confidence in Northern victory led to large financial losses for banks underwriting or holding U.S. bonds. On December 28th, 1861, the New York banks suspended specie payments. Other private banks followed suit immediately, and the federal government itself suspended specie payments on February 25, 1862 (Mitchell 1899; Bolles 1886, p. 68; Smith 1914; Friedman and Schwartz 1963). Unable to continue paying for goods and services in gold, the government turned to issuing irredeemable currency—Greenbacks.[15] The Greenbacks were crucial to the ratcheting up of federal expenditures and eventually a contentious part of the Civil War income tax repeal. Without Greenbacks, the Union would not have been able to fight such an expensive war (Dewey 1903).

Lincoln and the Congressional leadership knew that lack of enthusiasm for the war would hurt the Republicans in midterm elections. Some effort went into making sure that soldiers from key states were kept out of battle. Republican army officers observed their soldiers while voting, and federal troops oversaw elections in Border States (Anderson and Tollison 1991). Those maneuvers were important, but a finance system that did not offend key groups of voters was also important.

While the war was originally popular in most of the North, in other Northern states the war waxed and waned in popularity, and in yet other parts of the North it remained unpopular throughout. Generally, Northerners strongly supported the war, but some groups felt that the Civil War would mostly benefit the rich. The poor bore the burden of tariffs that were both high and regressive (McPherson 1988, p. 448). The poor who could not pay a bond to exempt themselves from military service would bear the burden of military duty. Although many of Irish descent enlisted in the Union army, voters in heavily Irish areas tended to doubt the war. New York State, for example, generally supported the war, but New York City did not.[16] New York State even had a serious anti-war candidate for governor.[17]

Using speeches by John Sherman (R-OH 13th District) and Justin Morrill (R-VT, 2nd District), Stanley (1993) argues that the poor in the industrial cities, Border States, and western agrarian states would have voted for taxes that were more

[14] On April 6 and 7, 1862 there were 23,746 estimated total casualties.

[15] Because they were expected to be (and were) eventually redeemed in gold, the greenbacks were technically a credit money and not a fiat money.

[16] Both New York City and Boston had draft riots.

[17] The anti-war movement was probably strongest in Ohio, but Ohio also contributed great numbers of soldiers. This appears to be one of the reasons that Senator John Sherman (R-OH, later Secretary of the Treasury) was such a staunch supporter of the income tax. Sherman felt that Ohio could do much for the Union if its people were staunchly behind it (Bensel 1991, p. 329; McPherson 1988, p. 560.)

progressive in nature, had they the option.[18] Politicians like Sherman and Morrill well understood the constraints they faced in building a winning coalition. They had to raise money to have finance capitalists willing to lend money. However, simply raising tariffs had limited potential for revenue and would anger the populace while an income tax could anger powerful constituents.

Furthermore, the wealth disparity was growing among the various sections of the United States. New York and New England were not only much richer than the western agrarian states, but they also had better access to capital. This income disparity led to some discontent that endangered Republican stability unless a concession could be included in the tax legislation.

The general consensus among writers, economists, and Congressmen was that tariffs, excise taxes, and the manufacturing tax could all be shifted to final consumers. The general public felt, however, that neither the income tax nor the property tax could be shifted (Stanley 1993; Ricardo 1823). Incomes were highest in New York and New England. Voters in western and Border States thus strongly favored an income tax. They opposed a property tax because they held the majority of their wealth in property.

Poorer voters in Border and other agrarian states, who supported the Union only marginally, did not have to pay the income tax because their income was too low. They did bear higher tariffs, which doubled, and excises on alcohol and tobacco, which increased dramatically. They also bore a disproportionate share of the regressive federal excise taxes on items like matches and playing cards. The progressiveness of the income tax, by contrast, made it very popular in agrarian districts (Taussig 1931).

The New England states were politically complex. Abolitionist support was strong in these states. Textile mill owners, small businessmen, and farmers strongly supported the war as well. But insofar as the states were industrialized, they were beginning to have large immigrant constituencies. These constituencies tended to vote Democratic, and the Democrats tended to oppose the war. The Northeast sector of American society owned about 70% of the nation's wealth in 1860. Naturally, it provided the most critical tax base, remitting 75% of the revenues.

Wealthy Northern states would pay most of the income tax and the manufacturing tax.[19] They also had relatively strong anti-slavery constituencies that comprised the bulk of the Radical Republicans (Bogue 1973). To some extent their Representatives were willing to accept an income tax in exchange for the end of slavery, even

[18] Stanley (1993) paints John Sherman as a centrist, but other writers including (McPherson 1988, p. 329 and 594) tell a story of Sherman as a staunch Republican. My view is that Sherman was a staunch Republican who was willing to make sacrifices to get the things he wanted. During the Civil War and Reconstruction about 75% of the Vermont vote went to the Republican Party. Congressmen like Morrill could certainly have taken very strong positions on a variety of issues, but as chair of the House Ways and Means Committee, Morrill was concerned about Republican fortunes in places beyond Vermont. At this time he was chair of the Subcommittee on Taxation.

[19] Luxuries were also taxed. These included gold plate, carriages valued over $50, yachts, and pool tables (Smith 1914).

though protective tariffs were preferred. The strongest promoters of the Civil War income tax were propertied Republicans like Justin Morrill, John Sherman, and Thaddeus Stevens. These politicians passed the Civil War income tax because other issues were more important than their immediate financial interests. Furthermore, the constituents who would pay the income tax were also strong supporters of the Civil War.

The upper middle classes of the nation's commercial and industrial centers complied widely with the income tax (Brownlee 2003). According to Stanley (1993), and Paul (1954), 10% of all Union households had paid the income tax by the war's end. Households in the northeast comprised 15% of that total. The North raised 21% of its war revenue through taxation, as opposed to the South, which raised just 5% this way.

Of course, voting patterns had to do with immigration patterns and ethnicity as well as industrialization. Southerners settled much of the lower Midwest while New Englanders settled the upper Midwest. States such as Illinois and Indiana were thoroughly split. Pennsylvania had pockets of anti-war sentiment, although it often took the form of anti-draft sentiment and anti-employer sentiment—the strikes in Pennsylvania's coal producing regions are an example (McPherson 1988; Mitchell 1899; Bogue 1973).[20]

With the defeat of the federal property tax in 1862, Congress increased income tax rates. The exemption decreased from $800 to $600, and all income between $600 and $10,000 was taxed at 3%. The rate on income over $10,000 climbed to 5%.[21] The 1862 bill was far more specific than the 1861 bill and established the Commissioner of Internal Revenue.[22] Still, the 1862 bill left many of the specifics up to the Treasury department.

4.2.5 Civil War Income Tax of 1864

The 1862 bill included both 1862 and 1863 because Congressmen almost seemed too busy to worry about the income tax. The 1864 bill, however, was far more contentious than the earlier bills. The national debt was already an astounding $1.8 billion, and the deficit was $600 million. This time Congress bitterly debated the degree of progressiveness. Salmon Chase argued that tax revenues should be enhanced to avoid the expense of borrowing even more money: "It is hardly too much, perhaps hardly enough, . . . to say that every dollar raised [by taxation] for extraordinary expenditures or reduction of debt is worth two in the increased value

[20] In the 1860 election, Pennsylvania sent 10 Democrats to Congress and 12 Republicans. Campbell of the 11th Congressional District (PA) was a Whig.

[21] The House Bill had a 7.5% tax on incomes over $50,000, but this was removed in the joint conference.

[22] George S. Boutwell was appointed the first Commissioner of Internal Revenue.

of national securities."[23] Wealthy states were well aware by this time that they were paying the lion's share of the tax. Massachusetts, with just over 6% of the North's total population, paid more than 13% of the income tax. New York paid more than 34% of the tax. These states argued against a highly progressive bill.

Furthermore, Congress also began to include all sorts of exceptions and special provisions that favored farmers in the income tax bill. For example, rent was deductible. Real estate income was taxed only if the property was bought and sold in the same tax year. Eastern interests and Midwestern manufacturers were willing to have an income tax, but they fought such efforts to shift the burden in their direction.

The June 1864 bill had a $600 exemption and taxed incomes between $600 and $5000 at 5%, income between $5000 and $10,000 was taxed at 7.5%, and income above $10,000 was taxed at the maximum rate of 10%. Interestingly, a month later the income tax rate was increased by 5% for all income above $600. Brownlee (2003, p. 34) estimates that by this time more than 10% of all Northern households paid the income tax and that in the northeast 15% of households paid the income tax. While tariffs and excises still provided the bulk of the tax revenue, income tax revenues were substantial by 1864. In 1866 the ratio of income tax to tariff revenue exceeded 40%. In 1865 and 1867, it exceeded 37%. In 1866 internal revenue collections reached $310 million; they were not that high again until 1911.

4.3 Civil War Income Tax Repeal, or What Happens After the Crisis?

In the standard ratchet model, taxes decrease quickly after the crisis but never to their pre-crisis level. The Civil War income tax is a twist on this. In the Peacock and Wiseman-Higgs model, voters have become accustomed to the higher taxes such that when the crisis is over, taxes never revert to their pre-crisis level. But the model does not explain why the income tax is completely removed after the Civil War.[24] It may still be partly correct because tariffs and excise taxes ratcheted, but it does not explain the income tax pattern.

The Civil War income tax and its repeal are better explained by the changing positions of various groups within the Republican Party. Since the Republicans were so powerful, one might wonder why they slowly abandoned the income tax instead of repealing it outright at the end of the war. Why did it take until 1872 to completely remove the income tax? Ratner (1967) argues that because the power of business interests increased slowly during this period, it took some time for them and their Republican allies to repeal the income tax. The Republican interest groups

[23] Report of the Secretary of the Treasury, 1863, p. 12. Originally found in Mitchell (1899). Chase also inflated the currency. Prior to the war, he was a hard money man. After the war he was again a hard money man, but during the war he was for inflation.

[24] Neither Peacock and Wiseman nor Higgs mention any outright tax repeals.

that formed the coalition to fight the Civil War and to raise the income tax were also slowly dissolving. Within the Republican Party, one group wanted to abandon the income tax immediately, one wanted to keep it, and one wanted to keep it for a short time.

Immediately after the Civil War the Republican Party became an amalgamation of several disparate groups whose views changed over time. The financiers or finance capitalists who lent the Union money during the Civil War were one group.[25] They were comfortable with a temporary income tax. The second group was comprised of manufacturers who were so concerned with the manufactures tax and with the tariff schedule that they were willing to accept tariffs on the high side of the Laffer curve. They saw the income tax as a substitute for tariffs and wanted to abandon the income tax immediately.[26] Yeoman farmers of the Midwest formed a third constituent group, one that became more active later in the period of Reconstruction and the Progressive Era and wanted to keep the income tax. A final group within the party consisted not so much of constituents but rather Congressmen who had personal agendas. These were the Radical Republicans,who wanted to punish the South and provide funding for the Freedmen's Bureau.[27] This group lost its identity relatively quickly after the Civil War.

Bensel (1991, pp. 301, 331, and 372) argues that the Republican Party was only able to keep its coalition of interest groups because it distributed wealth from the South to the Midwest. It also facilitated cross subsidies that allowed it to take on relatively unpopular policies including the gold standard and to some extent high tariffs.[28] As I show, the economic theory of regulation provides a better analytical framework for understanding the changes in U.S. tax policy than Higgs' voters get used to the higher taxes model.

[25]Banking interests are typically associated with New England and New York. These areas also represented traders opposed to the income tax because they were going to be paying it, yet favored the income tax because it might replace high tariffs, which weakened trade.

[26]Paul (1954) offers a slightly different perspective, arguing that banking and manufacturing interests were the force behind the movement to repeal the income tax, while merchants and importers were in favor of keeping the income tax and lowering tariffs. He does not split the country into the same three groups that I use.

[27]This is not to say that Radical Republicans' constituents were not also in favor of increased voting rights and punishing the South. However, it appears that during and immediately following the Civil War some Congressmen were elected who were more radical than the median voter in their districts.

[28]In some western states, Republicans split off from the main party because of the party's stance on gold. In some eastern states, some Democrats split off from the main party because of its stance on silver.

4.3.1 Manufacturers

Manufacturing interests were opposed to the income tax in part because factory owners paid the income tax, but more so because they preferred strong protectionist tariffs.[29] But they agreed to continue to vote for the income tax in exchange for votes for more tariffs from financial districts, until it appeared that the U.S. bonds were going to be paid off in gold. Further, prices were falling for manufactured goods (Sharkey 1959, p. 85). New England textile firms benefited from lower cotton prices, but other manufacturers were hurting. This made manufacturers eager to end the income tax and to pay for all government expenditures with a tariff.

4.3.2 Farmers

Most farmers did not pay the income tax, and those who did received many deductions.[30] Western farmers had little reason to favor the tariff, because it meant they paid higher prices for goods. Farmers argued that high tariffs reduced foreign demand for U.S. grain while an income tax would not.[31]

Further, the mild income tax helped to keep the illusion of vertical equity and to encourage the resumption of the gold standard. John Sherman (R-OH), for example, felt that a tax system based solely on regressive consumption taxes exacerbated class tensions among voters. This is a simplification, of course. Corn farmers were stronger anti-tariff constituents because corn tended not to be an export crop. Farmers west of Chicago tended to be more interested in railroads than tax policy. Farmers in the Midwest strongly opposed specie resumption (Atkinson and Beard 1911; Unger 1959). Many within the Democratic Party felt that government bonds ought to be paid for with Greenbacks. Their slogan was, "the same currency for the bondholder and the plowholder" (Garner et al. 1906, p. 1411). Men, particularly farmers, feared they would have to repay their debts in dearer currency than that in which they had contracted.

To coax western states to stay within the Republican coalition, they were paid off with farm-friendly legislation. The payoff to western farming states also included war pensions. Although the eastern states received more pensions, the wage differential made them much more important in the West. Republicans also paid off western states with railroad construction schemes and river and harbor

[29] Pennsylvania iron manufacturers were the most ardent advocates of this.

[30] The most contentious deduction was the fact that farmers who "ate" their profits did not pay taxes on them.

[31] Western Republicans eventually left the party. Progressives and Populists from the West sided with the Democratic Party (Bensel 1991; Galloway and Wise 1976).

improvements.[32] Southerners did not qualify for pensions, so this was a transfer from the tariff-paying South to the West. Congress made a protective tariff palatable to agrarian interests with the addition of a strong tariff on raw wool.[33] The Wool and Woolens Act of 1867 protected western farmers from British wool and to some extent wool from the Commonwealth countries (Taussig 1893).[34] Pennsylvania and Ohio were big wool producers, and their Congressmen pushed hard for the bill. Other farmers benefited as well, for many farmers throughout the United States kept a few sheep (Bensel 1991).[35]

4.3.3 Financial Interests

Historian Richard Bensel (1991) notes that the war had a huge impact on U.S. financial markets. They were fundamentally altered by the Union's tax and credit policies.

> The Union war mobilization probably fell most heavily upon the financial system, permanently changing both the internal organization of national capital markets and the relationship of the central state to finance capitalists (Bensel 1991, p. 238).

This group's preference for and against the income tax depended heavily on whether Greenbacks or gold would prevail. During the war British financial capital left U.S. markets, and there "emerged a distinctly American class of financiers" (Bensel 1991, p. 249).[36] Northeastern financial interests had the most complex interests, but, they were the major reason for the income tax's slow phase out instead

[32] During the election of the Speaker of the House in 1859, John Farnsworth (R IL) argued that if a Southern Democrat was elected to the Speakership, "the committees shall be so stocked and constituted that the rivers and harbors of the West shall have no protection whatever." *Congressional Globe*, 36:1:230, December 23, 1859. 87.7% of Free State Congressmen voted for the Rivers and Harbors bill in 1860, while only 20% of slave state Congressmen did so (Bensel 1991, p. 70).

[33] Democrats reminded Midwestern protectionists that other nations would match protectionism for protectionism. A tariff on wool would mean reduced markets for other agricultural products. For example, see the remarks made by Fernando Wood, Chairman of the Ways and Means Committee, on April 9, 1878.

[34] It should be mentioned that after the Civil War the income tax was not raising tremendous amounts of money. It was not the main card that Republicans were playing but rather one small piece of their win set. However, until other matters were taken care of, it could not be removed completely.

[35] In addition, a high tariff on lead appeased certain states that feared Mexican competition.

[36] "When the Civil War ended the interests of finance capitalists and the American state were probably more closely linked that at any other point in the nineteenth century" (Bensel 1991, p. 238).

of an abrupt end. Financial interests favored a temporary income tax in order to get specie resumption specifically for bond repayment.[37]

At the end of the Civil War, the United States owed $2,755,764,000 and was using Greenbacks instead of gold (US Bureau of the Census 1976, p. 1118; Bolles 1886). Finance capitalists were primarily concerned with the return to the gold standard and the repayment of government bonds in gold.[38] Since the wartime boom ended in April 1865, financiers worried that a post-war recession would prevent the repayment of bonds in gold (Bolles 1886). The income tax, which they paid, was of secondary importance. Finance capitalists were willing to make sacrifices in order to fill federal reserves with gold and to hold off western pressure for soft money and easy credit.

Greenbacks were discounted heavily against gold lowering the value of bonds which might be repaid in Greenbacks. The discount also made international transactions more costly. All foreign transactions took place in gold while domestic transactions were carried out in Greenbacks. Hence Greenbacks provided both currency risk and transactions costs to anyone who wanted to trade in either foreign goods or foreign capital markets. As Friedman and Schwartz (1963, p. 26) note:

> Dealers as well as others having extensive foreign transactions therefore found it convenient to maintain gold balances as well as greenback balances. To accommodate them, New York banks and perhaps others as well, had two types of deposit accounts: the usual deposits payable in greenbacks or their equivalent and special deposits payable in gold. The gold deposits were expressed in dollars' like the greenback deposits but that dollar meant a very different thing. It stood for the physical amount of gold that had corresponded to a dollar before the Civil War and was to again after 1879. During the period of suspension, this physical amount of gold was worth well over two dollars in greenbacks from mid 1864 to early 1865..

The financial class saw the Greenback discount as their foremost problem. They wanted to add gold to the federal reserves which made them favor quick recovery of cotton production and export. Anything that slowed gold accumulation was considered an anathema. Many in Congress hoped that as the economy grew, it would grow into the money supply and the Greenback discount disappeared. At that point the U.S. could resume using gold for currency with limited pain. When

[37]The northeast includes not only finance capitalists but also manufacturers and international traders. Following Bensel (1991), at least some districts finance capitalists were extraordinarily strong. Furthermore, while international traders faced increased risk when they had to purchase their goods in gold but sell them for fiat money. While they could hedge their risk by buying options, they greatly preferred the gold standard. Moreover, chambers of commerce regularly lobbied the federal government for an end to the Greenback.

[38]It might be argued that specie resumption was a payoff for the risky loans that finance capitalists took during the Civil War, but a more likely explanation points to their relative power. Mayer (1964) suggests a close association between Northeast financiers and a wide variety of Republican interests, including state parties in Ohio and Indiana. Mayer cites one Republican manager from Indiana who sent a letter to Jay Gould asking for $10,000 in 1876 to spend getting Indiana to vote hard money and Republican. Mayer also cites the close association with financiers when it came to defeating the Greenback party in the Midwest (Mayer 1964, p. 199).

the business cycle downturn ended at the end of 1867, increased money demand lowered the Greenback discount (Sharkey 1959, pp. 107, 117).

The business cycle upturn also provided more reason to believe that the currency would be paid off. This provided increased impetus for Congressmen from the Northeast to vote to lower the income tax. By 1880 the federal debt had been cut from $2.7 to $2 billion. Per capita debt had been cut in half. Because of the finance capitalists, all proposals for increased taxes on government bonds were defeated in the 1860s.

Had it not been for a recession in 1872, the United States likely would have returned to gold shortly after the end of the income tax. As post-Civil War expenses wound down and the federal government slowly improved its chances of returning to gold, raising the exemption and reducing the tax rate lowered the income tax overall. The falling price of gold—from $2.019 in 1865 to $1.233 in 1870 to $1.120 in 1872—allayed fears that there would never be specie resumption (Mitchell 1908, pp. 5–13). Once it looked like the U.S. would return to the gold standard, keeping taxes seemed much less important. Finance capitalists and associated bondholders no longer fought for an income tax, but manufacturing groups did fight.[39] Since financial interests were paying the income tax, they had reason to fight it after their primary concern was dealt with.

4.4 Interaction of the Three Groups

Because one group wanted an income tax, one did not, and the third group was ambivalent, a priori thinking would imply that a small income tax would result. This is especially true given John Sherman's feelings that a small income tax was stabilizing for the economy. However, the Northeast and the northern industrial belt held the most important positions within the Republican Party, and their interests dominated. Western interests were poorly represented, and they were paid off with a wool tariff and increasingly with Union pensions. Additionally, the South could be made to pay taxes through the tariff and the cotton excise, but virtually no Southerners had enough income to pay an income tax. The only Southern cities with substantial income tax receipts were Knoxville, TN, and New Orleans, LA.

When Andrew Johnson succeeded Lincoln, he was all but powerless. The defeated South was under military rule of the North and relatively powerless even after the Hayes-Tilden election. The Republican Congress held all the real power.[40] Republicans found that in many districts they could wave the bloody shirt, either

[39] Kindahl argues that species resumption would not have occurred if there hadn't been major deflation as a result of currency contraction and recessions.

[40] The "American state and the Republican Party were essentially the same" from 1861 to 1877 (Bensel 1991, p. 3). At the end of the Civil War, the Republican Party leadership consisted mostly of wealthy Northeasterners (Bensel 1991).

figuratively or actually, and win the election. In contrast, the Democratic Party was both strongly associated with Southern disloyalty and split between soft money and hard money interests.[41]

On the expenditure side, Reconstruction was taking a major toll on the finances of the United States. Effectively over by 1871 with Hayes-Tilden, Reconstruction had been expensive, and transfer of all the debt into a tariff was impossible. The total number of armed forces was 150% of the pre-Civil War number. Federal military expenditures increased and veterans' benefits were increasing, although they had yet not reached the level they would in the 1880s.[42] Congress also found other areas on which to spend money; for instance, it began financing four railroads across the continent. Thus, only after Reconstruction could manufacturing interests vote to get rid of the income tax.

4.5 Conclusion: Lessons From the Civil War Income Taxes

> The secession of the South in effect broke the logjam behind which this (Northern Republican) agenda had languished in the years just prior to the Civil War and a major portion of state expansion was composed of policies that had been proposed and debated in the prewar period. (Bensel 1991, p. 2).

At the outset of the Civil War, Northern constituents could see a clear need for an increase in revenue. Northern voters were willing to accept higher taxation, but more importantly the traditional low-tariff proponents, Southerners, were absent from the debates. This allowed Union interest groups to form a winning coalition that existed in a very different part of the relevant policy space. Immediately after the war, Southerners and Copperhead Democrats were powerless. Farmers, manufacturers, and bankers struggled to find an appealing tax policy. This suggests that the traditional Peacock and Wiseman-Higgs theory of the ratchet does not hold. Changing taxpayer willingness is not the most important element in the ratchet. Rather, crises break up previous equilibria and give an advantage to certain groups when the crises end.

While the literature on the ratchet effect suggests that ratchets are voter driven, anecdotal and empirical works suggest that voters are just one of many interest groups. Voters in different areas certainly had different views. Simply positing that all voters change their ideology on the appropriate level of taxes is incorrect.

This case study has shown how a specific aspect of government growth can occur and how it need not be permanent at all. The ratchet effect only works when the interest group favoring the increase is either more powerful than those opposed to the growth of government or can successfully utilize the status quo bias. When

[41] Interestingly, the Republicans and Democrats switched their stances on strong currency during the Civil War.

[42] By 1884 veterans benefits would account for 29% of all federal expenditures.

Rasler and Thompson (1985) looked at the ratchet effect, they found that a ratchet usually occurs but not always. This paper suggests that looking at the interest groups formed during the crisis is the best way to understand when a ratchet effect might occur.[43] The coalition of interest groups and the relative power of each matter greatly when studying the ratchet.

References

Anderson GM, Tollison RD (1991) Political influence on civil war mortality rates: the electoral college as a battlefield. Def Peace Econ 2(3):219–233
Atkinson CR, Beard CA (1911) The syndication of the speakership. Political Sci Q 26(3):381–414
Bensel RF (1991) Yankee leviathan. Cambridge University Press, Cambridge
Bogue AG (1973) Radical voting dimensions. J Interdiscip Hist 3(3):449–474
Bolles AS (1886) The financial history of the United States, vol 1–3. D. Appleton, New York
Brownlee WE (2003) Funding the modern American State, 1941–1995: the rise and fall of the era of easy finance. Cambridge University Press, Cambridge
Dewey DR (1903) Financial history of the United States. Longmans, Green, and Company, New York
Friedman M, Schwartz AJ (1963) A monetary history of the United States, 1867–1960. Princeton University Press, Princeton
Galloway GB, Wise S (1976) History of the house of representatives. Thomas Y. Crowell Company, New York
Garner JW, McMaster JB, Burdick JE (1906) The history of the United States, vol 4. John D. Morris and Company, Philadelphia
Higgs R (1987) Crisis and leviathan. Oxford University Press, New York
Holcombe RG (1999) Veterans interests and the transition to government growth: 1870–1915. Public Choice 99(3):311–326
Johnston L, Williamson SH (2003) The annual real and nominal GDP for the United States, 1789–2002. Econ Hist Serv. http://wwwehnet/hmit/gdp
Kinsman DO (1903) The income tax in the commonwealths of the United States. Publ Am Econ Assoc 4(4):1–128
Mayer GH (1964) The republican party, 1854–1964. Oxford University Press, Oxford
McPherson JM (1988) Battle cry of freedom. Oxford University Press, Oxford
Mitchell WC (1899) The suspension of specie payments, December 1861. J Polit Econ 7(3):289–326
Mitchell WC (1908) Gold, prices, and wages under the greenback standard. University of California Press, Berkeley
Olson M (1965) The logic of collective action. Harvard University Press, Cambridge
Paul RE (1954) Taxation in the United States. Little, Brown and Company, Boston
Peacock AT, Wiseman J (1961) The growth of public expenditure in the United Kingdom. Allen & Unwin, London
Rasler KA, Thompson WR (1985) War making and state making. Am Polit Sci Rev 79(2):491–507

[43]Of course, the famous Spanish-American War telephone tax that lasted for one hundred years, but that is a different question. In 1898 the U.S. instituted a telephone tax to pay for the Spanish-American War. That tax was repealed in 1998. Since no organized constituent group opposed the small tax, it lasted. The Civil War income tax was small compared to some other taxes, but organized groups wanted it removed.

Ratner S (1967) Taxation and democracy in America. Wiley, New York
Ricardo D (1823) Principles of political economy and taxation. E.P. Dutton & Co., Inc, New York
Seligman ER (1895) The income tax in the American colonies and states. Polit Sci Q 2:221–247
Sharkey RP (1959) Money, class, and party. Johns Hopkins Press, Baltimore
Smith HE (1914) The United States federal internal tax history from 1861 to 1871. Houghton Mifflin Company, New York
Stanley R (1993) Dimensions of law in the service of order. Oxford University Press, Oxford
Stevens T (1997) The selected papers of Thaddeus Stevens, volume 1: April 1865-August 1868. University of Pittsburgh Press, Pittsburgh
Tabarrok A, Spector L (1999) Would the Borda count have avoided the civil war? J Theor Politics 11(2):261–288
Taussig FW (1893) The duties on wool and woollens. Q J Econ 8:1–39
Taussig FW (1931) The tariff history of the United States. A.M. Kelley, New York
Unger I (1959) Business men and specie resumption. Polit Sci Q 74(1):46–70
US Bureau of the Census (1976) The statistical history of the United States, from colonial times to the present. Government Printing Office, Washington, DC
Witte JF (1985) The politics and development of the federal income tax. University of Wisconsin Press, Madison

Chapter 5
Who Did Protective Legislation Protect? Evidence from 1880

Jeremy Atack and Fred Bateman

Abstract Beginning in the 1840s many states passed laws mandating the compulsory education of children and regulating the work of women and children although these were far from universal by 1880. In this paper, we focus on the impact of hours laws, especially those for women. Scholars have raised serious questions about the effectiveness of these laws because of doubts about enforcement mechanisms and whether or not the laws were binding. Moreover, it has been questioned as to whether these laws were simply passed as part of rent-seeking behavior by those not covered by the laws, in particular, adult men. In response, many of the laws covering adult women have now been rolled back. One state, Massachusetts, however, did pass an effective law in 1874 that resulted in the (successful) prosecution of at least one politically powerful corporation. Here, we investigate the impact of these laws using establishment level data for 1880. The historical record is consistent with rent-seeking by men but not for the purpose of disadvantaging women. The historical record is consistent with rent-seeking by men but not for the purpose of disadvantaging women. Rather, men pressed the case for women and children to secure benefits that they were apparently unable to achieve on their own. This was possible because, at the time, women and children were complements to male labor rather than substitutes. We find that there were systematic variations in hours from industry to industry, between city and countryside and regionally and that violations of the laws was not uncommon. Larger firms such as those in urban areas or those employing large numbers of the affected group were, however, more likely to be in compliance, particularly in Massachusetts. The evidence for Massachusetts also suggests, albeit very weakly, that the magnitude and certainty of penalties for violating the law may have been a major factor determining compliance with the law.

J. Atack (✉)
Vanderbilt University, Nashville, TN, USA
e-mail: jeremy.atack@vanderbilt.edu

F. Bateman, deceased (1937–2012)
University of Georgia, Athens, GA, USA

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_5

5.1 Introduction

> Ten hours' labor per day has been the recognized work-day in nearly every mechanical and manufacturing industry in the United States, for the past thirty or forty years, except in the textile manufactories of the eastern States; and in them, as in England, *legislation has been invoked specially for women and children, and which, being secured, caused a lessening of the hours of adult male labor* … (Ohio Bureau of Labor Statistics 1879, p. 266; emphasis added)

> Although some supporters of this legislation [for women and children] apparently hoped that it would serve as a stimulus to similar policies for adult males, it failed to have this effect. The achievement of shorter hours for men was to come mainly through collective bargaining. (Derber 1962, p. 273)

In 1874, after decades of industrial action, political agitation and lobbying by interested groups, the Massachusetts state legislature finally passed a law that became the model for similar legislation in other states. That law established statutory limitations upon the hours that women could work by prohibiting the employment of women for more than 10 h a day except to make up time lost by breakdowns or to secure a shorter working day sometime during the week and in no case were their hours to exceed 60 in any 1 week.[1] In addition, the Massachusetts law forbade employees from contracting out of the provisions either at their own or their employers behest and all existing labor contracts calling for longer hours were invalidated.

Surprisingly, this law also withstood legal challenges. In the case of *COMMONWEALTH V. HAMILTON MANUFACTURING COMPANY*, the Massachusetts Supreme Court ruled that:

> [The 1874 law] merely provides … no person shall be engaged in labor more than ten hours a day or sixty hours a week. There can be no doubt that such legislation be maintained either as a health or police regulation … This principle has been so frequently recognized in this Commonwealth that reference to the decisions is unnecessary.[2]

Although this decision was not appealed to the U.S. Supreme Court, we would argue that the Massachusetts law broke with the constitutional rights guaranteed by Article I, Section 10 and affirmed in the *DARTMOUTH COLLEGE V. WOODWARD* decision that declared "no state shall pass any … law impairing the obligation of contracts."[3] Further, we believe that the successful prosecution of one of the largest manufacturing firms in America at the time sent a clear message that such laws could and would be enforced.[4] The Massachusetts law, however, was to remain the exception rather than the rule until the late 1880s when Maine passed a similar

[1] Massachusetts, Laws 1874, C. 221.

[2] COMMONWEALTH V. HAMILTON MANUFACTURING COMPANY, 120 Mass. 383 (1876).

[3] *DARTMOUTH COLLEGE, TRUSTEES OF V. WOODWARD* 17 US 518, 4 L ed. 629 (1819).

[4] At the 1870 Census of Manufactures, the Hamilton Manufacturing Company was reported to have a capital of $1.2 million and employed 350 men, 672 women and 32 children to produce more than 10 million yards of cloth valued at $1.3 million. It was among the largest manufacturing firms in

law. Even by 1896, only 13 states had passed any kind of law restricting women's hours and of these one had been declared unconstitutional by the state supreme court.[5] Only Maine, Massachusetts and New Jersey had laws thought to be effective (Brandeis 1918, p.457).

In 1908, however, the U.S. Supreme Court finally settled the question of constitutionality in *MULLER v. OREGON* affirming that such laws were reasonable uses of police power.[6] Consequently, by 1921, all but four states had adopted such laws (Commons and Andrews 1927, p. 249). The route to success, as with the Massachusetts law, lay in the appeal to public health as the defense for this use of police powers.

The Massachusetts legislation did not evolve in a vacuum. Rather it was a part of a wider movement, international in scope, to protect not only specific groups—especially women and children—but also labor in industry in general. For women, such laws included prohibitions against night-work[7] and required seats for women (but not men) at work stations.[8] These laws were also part of an even larger corpus expanding governmental interference in American economic life that began in the latter half of the nineteenth century, traditionally dated from *MUNN V. ILLINOIS*.[9] Markets were seen as imperfect and in need of regulation to achieve newly articulated economic and social goals.

The wheel has now almost turned full circle and many of those protections have since disappeared. The erosion of protective legislation began with an August 19, 1969 ruling by the Equal Employment Opportunity Commission that "such laws and regulations conflict with Title VII of the Civil Rights Act of 1964 and will not be considered a defense to an otherwise established unlawful employment practice."[10] As a result, "state legislatures again made adjustments in State maximum hours laws making more job opportunities available to women" (Mitchell and Sorenson 1970, p. 54). For example, women in New York state on issuance of a state labor commissioner permit were once again permitted to work beyond the 8 h a day and

---

the U.S. See the manuscripts of the Census of Manufactures at the Ninth Census for Massachusetts, p. 127, line 8, held by the Massachusetts state archives in Boston.

[5] *RICHIE v. PEOPLE*. 155 Ill. 98 (1895). In 1909, Illinois adopted a new ten-hour to replace the eight-hour law overturned in *RICHIE v. PEOPLE*. Its constitutionality was almost immediately challenged and upheld in *RITCHIE V. WAYMAN*, 244 Ill. 509, 91 N.E. 695 (1910) using the new secret weapon developed for *MULLER V. OREGON*—the Brandeis brief.

[6] Regarding the importance of police powers, see US Department of Labor (1904). At least four state courts had sustained such legislation: *COMMONWEALTH V. HAMILTON MANUFACTURING COMPANY*, 125 Mass. 383; *WENHAM V. STATE*, 65 Neb. 394, 400, 406; *STATE V. BUCHANAN*, 29 Wash. 602; *COMMONWEALTH V. BEATTY*, 15 Pa. Sup. Ct. 5, 17. For the U.S. Supreme Court decision, see *MULLER V. OREGON*, 208 US 412.

[7] For example, Massachusetts, Laws 1890, C. 183. See also Commons (1918, p. 288–94).

[8] Commons and Andrews (1927, p. 401–03) claim that New York was the pioneer in this field, passing a law as early as 1881, but without reference to the specific statute.

[9] *MUNN V. ILLINOIS*, 94 US 125 (1877).

[10] Federal Register, 34, no.158, August 19, 1969.

48 h a week standard, returning rights denied them since 1899. In addition, women aged 18–21 could work until midnight and women over 21 could even work later (Mitchell and Sorenson 1970).

More significantly, in March 1990, the U.S. Supreme Court accepted a case on appeal that struck at the very heart of protective labor legislation covering women.[11] In this case, Johnson Controls, a maker of lead-acid automobile batteries, sought to maintain its fetal protection policy.[12] The company policy, adopted in 1982, prohibited the employment of fertile women in jobs that brought them into contact with the lead dust. The justification for the policy was to absolve Johnson Controls from any potential liability for birth defects that may be caused by—or blamed upon—exposure to high concentrations of airborne particles of lead even when those lead levels remained within generally accepted safe limits for exposure adopted by the Environmental Protection Agency and the Occupational Safety and Health Administration. Two years after adoption, the United Auto Workers union filed two class action suits on behalf of its members: One alleged that the company policy of transferring women to other jobs where they would not be exposed to lead dust harmed the chances of men to transfer to those job; the other claimed that women were denied equal access to the high paying—and therefore desirable—jobs where they would be exposed to lead dust.

An initial decision favoring Johnson Controls was rendered in 1988 by U.S. District Court for the Eastern District of Wisconsin Judge Robert Warren,[13] and affirmed by the 7th U.S. Circuit Court of Appeals.[14] The U.S. Supreme Court, however, in a unanimous ruling issued after this paper was originally written, overturned the U.S. District Court and Appeals Court decisions.[15] In the Court's opinion Johnson Control Inc.'s "fetal protection" policy violated Title VII of the Equal Rights Act of 1978 amended to include the Pregnancy Discrimination Act: "Johnson Controls' professed moral and ethical concerns about the welfare of the next generation do not suffice to establish a bona fide occupational qualification."[16]

[11] *INTERNATIONAL UNION, UAW V. JOHNSON CONTROLS, INC*. This case was accepted by the U.S. Supreme Court, March 26, 1990. See *Los Angeles Times*, March 27, 1990, A1:3 and A20:1.

[12] The following facts of the case have been taken from 886 F.2d 871 (7th Cir. 1989), that is West's Federal Reporter, Cases Argued and Determined in the United States Courts of Appeal and Temporary Emergency Court of Appeal, (St. Paul: West Publishing Co., 1990), Volume 886 F.2d, pp. 871–921.

[13] 680 F.Supp. 309.

[14] 886 F.2d 871 (7th Cir. 1989).

[15] *New York Times*, March 21, 1991, A12. For the Court's full decision, see 499 U.S. 187 (1991).

[16] The Supreme Court decision closely mirrors a prior decision against Johnson Controls by Justice Robert Staniforth of the state appeals court in Orange County, California who determined "however laudable the concern by business . . . for the safety of the unborn, they may not effectuate their goals . . . at the expense of the woman's ability to obtain work for which she is otherwise qualified." See "Company loses 'fetus protection' case," *Orange County Register* (Santa Ana, California), March 2, 1990.

Furthermore, the Justices argued that "decisions about the welfare of future children must be left to the parents who conceive, bear, support, and raise them."[17]

Women, however, were not the only groups protected from the marketplace. Children were, too, and from an even earlier date. Beginning in the 1830s, some states prohibited the employment of children below a certain age. Others regulated the hours and conditions under which children could work. Effective federal laws regulating child labor date from the Fair Labor Standards Act of 1937. In addition, compulsory education laws were passed ostensibly to ensure that children had some minimum opportunity to consume a social good. Sometimes these compulsory education laws were complementary with other child labor laws, but they may have been substitutes for other forms of regulation of child labor. These laws are not currently under attack. Indeed, newspaper stories suggest that the U.S. Department of Labor has strengthened its enforcement of child labor laws in recent years.

In this paper, we trace the origins of protective legislation focusing upon laws regulating hours of work and examine the quantitative and qualitative evidence on the impact of these laws. Our analysis raises many questions to which we seek answers: Whom did protective legislation protect? Were women and children truly protected by laws limiting their hours of work and, if so, were they the primary beneficiaries? Did these laws also benefit workers who were not covered by the regulations? Did legislation limit the employment opportunities of those affected? Did it lead to the concentration of protected workers in specific fields—the "feminization" of certain occupations? If the laws were ineffective, how widespread were violations? How important was the means of enforcement for its effectiveness?

Unfortunately, our results do not resolve all of these questions. Our quantitative data are just for 1 year, 1880, a time when only five states had adopted legislation specifically setting limits on women's hours. However, another twelve states had also passed hours laws covering all workers, 12 had passed limits on children's hours and 14 had adopted compulsory education for children to age 14 (and in some cases, for even older children). Further, many of the issues are not simple "either-or" questions. For example, men may have benefited from legislation protecting women and children without diminishing the gains to those protected (as the header quote from the Ohio Bureau of Labor Statistics suggests). Moreover, everyone may have gained—or lost—depending upon the complementarity or substitutability between men, women, and children and between labor and other factors of production. Thus, behavior that is consistent with the tenets of "rent seeking" by male employees—

[17] By failing to acknowledge the state's interest in the welfare of children, the decision in Johnson Controls also appears to reverse the policy enunciated in *MULLER v. OREGON* (1908) that: "healthy mothers are essential to vigorous offspring [so that] the physical well-being of woman becomes an object of public interest and care in order to preserve the strength and vigor of the race . . . differentiated by these matters from the other sex, she is properly placed in a class by herself, and legislation designed for her protection may be sustained, even when like legislation is not necessary for men and could not be sustained." See *MULLER v. OREGON*, 208 US 412, pp. 6–7.

reduced job opportunities for women increasing demand for male labor—may be motivated by altruism as much as misanthropy and be self-serving while simultaneously also serving others.

5.2 The 1880 Census Sample

The quantitative data underlying this study are taken from a random sample of over 8000 firms drawn from the 1880 Census of Manufactures.[18] These data, for most firms, include information on the employment of men, women, and children, their hours of work between May and November and between November and May, plant location, capital invested, the value of inputs and outputs, and their source of power.[19] Table 5.1 shows the 2-digit SIC Code and number of firms and employees for each industry.

Although firms in the food processing (SIC 20) and lumber milling (SIC 24) industries were the most common, there were substantial numbers of firms and employees in most of the industries. Collectively, firms in the nationally representative sample from the 1880 Census employed almost 87,000 workers. The sample statistics generally mirror those for the nation as a whole, except for those industries—including textiles, a major employer of women—that were collected by specialist agents whose returns were not included among the 1880 census manuscripts.

Of particular interest for our analysis are the data on hours of work. The reported summer (May-November) and winter (November-May) hours are assumed to be scheduled, not actual, hours of work because of the almost uniform reporting of whole numbers of hours (Atack and Bateman 1992, 13–16). In order the generate a best estimate of the length of the scheduled working day we averaged winter and summer hours, except for 32 firms that reported zero hours for each 6 month period.[20] If a firm reported no hours for just one of the 6-month periods, we used the reported work hours for the non-zero 6-month period as the average number of

[18]These data were collected by Jeremy Atack and Fred Bateman with funding by the National Science Foundation to the University of Illinois under SES 86-05637 and Indiana University under SES 86-09392. They are available from ICPSR at the University of Michigan as ICPSR 9384 and 9385 (Atack and Bateman 2004a,b); They are available electronically at https://doi.org/10.3886/ICPSR09385.v2 https://doi.org/10.3886/ICPSR09384.v2. Complete documentation is available at the ICPSR and readers are also referred to Atack and Bateman (1999, p. 177–88).

[19]See Wright (1900, p. 315–61) for a complete list of the census questions including those on the special schedules.

[20]In addition, there were 136 establishments in which summer hours were missing (as distinct from zero), 135 with missing winter hours and 141 firms with both summer and winter hours missing. These firms were excluded from the calculations.

Table 5.1 The industrial distribution of firms and employment, by 2-digit SIC industry, in the 1880 Atack-Bateman sample

Industry	2-Digit SIC code	Number of firms	Number of employees
Agricultural services	7	56	249
Construction	17	685	5776
Food	20	1440	7723
Tobacco	21	239	4299
Textiles	22	47	789
Clothing	23	316	14,647
Lumber	24	1224	8878
Furniture	25	210	2588
Paper	26	78	2045
Printing	27	140	2845
Chemicals	28	178	3044
Leather	31	992	7739
Glass/brick	32	212	2503
Primary metals	33	306	4136
Fabricated metals	34	81	3591
Machinery	35	444	6996
Non-ferrous metals	36	3	305
Trasport. equipment	37	46	2322
Instruments	38	8	70
Misc. metals	39	228	3060
Coke/gas/oil	49	28	1952
Blacksmithing	76	786	1390
All industries		7747	86,947

Source: Data from Atack and Bateman (2004a)

hours of work per day.[21] Given the widespread opposition to long hours of work, we have also imposed an arbitrary limit of 14 h on the length of the scheduled working day. For the small number of firms reporting longer hours than this—just over 1% of all firms that reported summer hours and well under 1% of firms reporting winter hours—we have arbitrarily assumed that they employed two shifts, each working one-half of the reported hours per day.[22]

Data in the published census volumes allow calculation of the fractions of industry employment represented by adult male and adult female workers and by children and their distribution across states. The industrial labor force numbered

[21]Forty-nine firms (0.6% of the sample) reported zero hours for one of the 6 month periods. Of these, 31 reported not operating in winter and were typically firms in flour milling, lumber, and brickmaking. Twelve of the 18 firms that reported not working in summer were in the agricultural services industry.

[22]That is to say, firms reporting 15 h of work were recorded for these calculations as operating 7.5 h per worker, but 15 h per firm.

2,738,892, about 20% of whom were adult females and 6.6% were children. New York had the largest number of female workers, but women made up a larger fraction of the labor force in Massachusetts and New Hampshire—about a third. In Montana, on the other hand, only three women were employed in industry and they represented less than one half of 1% of the industrial labor force (U.S. Department of the Interior, Census Office 1883, p. 928).

Shares in the sample were similar (Table 5.2). Adult females made up almost 20% of the labor force and about 5% were children. We have used the firm-level data in the sample to estimate two other useful sets of statistics. The first of these is the average proportion of each firm's labor force that was female or child by industry. This figure will be equal to that for the industry as a whole if each firm in the industry employed the same proportion of women and children. If, on the other hand, large firms employed disproportionately more women, then the average across firms will be less than average for the industry as a whole. The second reports the fraction of firms in each industry that employed any women or children. This measures how widespread the employment of women and children was in a particular industry.

Table 5.2 reveals that adult females made up only 4.7% of the labor force of the average firm while children constituted only 2.7% of the average firm's workforce. Nevertheless, women made up almost 20% of the industrial labor force and children, 5%. The inconsistencies between these figures reflect—as others have also found—the tendency for women and children to be more extensively employed by the larger firms and in specific industries (Goldin and Sokoloff 1982, pp. 741–74, especially Table 2 and pp. 751–756). Consider, for example, the textile industry (SIC 22). Across our sample firms in this industry women represented only 18.6% of the labor force in each firm, yet at least 49.2% of all workers in this industry were women.[23] This greater reliance upon women and children in larger firms is attributable to the greater division of labor in these firms that resulted in the de-skilling of jobs and the substitution of mechanical effort for human muscle.

The data also reveal the concentration of women and children in certain industries. Women made up at least 20% of the labor force in tobacco, textiles, clothing, paper, and miscellaneous metals and in another four industries—printing, chemicals, leather, and fabricated, non-ferrous metals—women made up at least 10% of the workforce. Children made up at least 10% of the labor force in the agricultural services industry, tobacco, textiles, furniture and paper. We will refer, somewhat arbitrarily and loosely, to those industries where the labor force was at least 20% female as "female-dominated" industries and those where children were 10% or more of the workforce as "child-dominated" industries. Semantics aside, these are clearly industries in which women and children were important and in which they tended to be important regardless of firm size. At the opposite extreme,

[23] As noted above, this latter statistic is biased downward in our sample because of the separate enumeration of the large mills in the textile centers of New England—Lowell, Lawrence and Fall River—and their exclusion from our sample. For example, the published census yields an estimate of 57% female in the cotton textile industry. See U.S. Department of the Interior, Census Office 1883, p. 1125

Table 5.2 The extent of employment of women and children in the 1880 census sample by industry (percentages)

	Composition of labor force, % by industry[a]		Composition of labor force, % by firm[b]		% of Firms employing[c]		
Industry	Women	Children	Women	Children	Women	Children	Either
Agricultural services	2.0	10.8	1.5	13.1	5.0	23.3	28.3
Construction	0.8	2.6	1.0	2.4	1.8	8.9	10.1
Food	7.1	4.5	2.6	2.2	6.5	6.0	10.8
Tobacco	26.1	12.4	6.7	6.8	20.2	20.2	30.7
Textiles	49.2	11.4	18.6	5.6	40.8	22.4	49.0
Clothing	68.2	2.6	50.2	2.4	67.7	11.7	69.8
Lumber	0.9	3.3	0.4	1.8	2.3	7.0	8.9
Furniture	9.4	15.5	3.1	4.4	10.6	14.2	19.5
Paper	27.9	11.5	24.6	8.1	43.6	32.1	61.5
Printing	14.9	9.3	12.5	8.6	41.8	34.2	56.2
Chemicals	17.2	5.8	9.1	5.4	24.5	15.2	31.0
Leather	15.6	4.6	3.4	1.6	9.5	5.4	12.3
Glass/brick	3.2	7.6	1.9	8.7	7.5	29.7	33.5
Primary metals	4.0	03.6	1.2	2.0	3.7	7.5	10.0
Fabricated metals	10.6	7.4	5.0	3.7	17.3	18.5	25.9
Machinery	0.7	3.3	0.3	2.1	1.3	9.3	9.9
Non-ferrous metals	10.5	8.2	25.6	2.8	66.7	33.3	66.7
Transport equipment	0.2	2.9	1.6	2.3	4.3	8.7	10.9
Instruments	0.0	1.4	0.0	2.1	0.0	12.5	12.5
Misc. metals	23.6	6.1	8.8	5.7	19.5	16.3	31.7
Coke/gas/oil	0.5	0.2	0.0	1.1	3.6	3.6	7.1
Blacksmithing	0.5	0.6	0.2	0.5	0.6	1.0	1.5
All industries	19.1	5.0	4.7	2.7	10.0	9.2	16.3

Source: Data from Atack and Bateman (2004a)

[a] % of women or children employed in industry = $\left[\sum_i (Women_i \text{ or } Children_i) / \sum_i Men_i + Women_i + Children_i)\right]$*100, for each j where i = 1, . . . , n firms in industry j

[b] Average percentage of women or children employed by firms in industry j = $\left[\left[\sum_i (Women_i \text{ or } Children_i) / \sum_i Men_i + Women_i + Children_i)\right]/\text{n}\right]$*100, for each j where i = 1, . . . , n firms in industry j

[c] Percentage of firms employing women or children = [(number of firms in industry j employing women or children)/(number of firms in industry j)]*100

very few women (under 1% of the labor force) were to be found in construction, lumber, machinery, transportation equipment, instruments, coke, gas and oil, or blacksmithing.

The last columns of Table 5.2 provide some evidence on the number of firms that could be affected by restrictions on the employment of women and children. Some firms in virtually every industry employed women and/or children. Overall, about 10% of all firms employed at least one adult female and more than 9% employed

one or more children, with 16% employing one or the other or both. The potential constituency that could be affected by protective legislation was thus quite large. Moreover, in some industries—such as textiles, clothing, paper and printing, where 40% or more of the firms employed at least one woman and paper and printing where perhaps a third employed children—we would expect to see organized opposition if legal restrictions in employment adversely affected profitability.

5.3 Employee Opposition to Long Hours and the Length of the Working Day

One obvious question in the context of this paper is whether workers (male or female) wanted shorter hours. Unfortunately, the historical evidence is somewhat mixed before the 1880s. Demands for shorter hours were among the earliest industrial disputes in the new nation. For example, in 1791 the Philadelphia Carpenters struck for shorter hours and overtime pay (Commons 1918, p. 69). Over the succeeding decades, Philadelphia was to remain the center of agitation for shorter hours. In 1827, for example, the building trades struck for a ten-hour day. That strike also failed but out of it emerged the first effective centralized city organization of wage-earners in the world—the Mechanics' Union of Trade Associations—which in turn gave birth to the first labor party—the Working Men's party—that led in turn to the first industrial trade union—the New England Association of Farmers, Mechanics, and other Workingmen (Commons 1918, p. 169). Beyond trying to secure relief from long hours of work for individual workers, many labor groups saw shorter workdays and shorter work-weeks as the solution to unemployment and the means for increasing industrial employment generally.[24]

The Ohio Commissioner of Labor alleged that reducing the hours of work also figured prominently at all labor gatherings after the Civil War (Ohio Bureau of Labor Statistics 1879, p. 259). However, an analysis of the 762 strikes in the 1880 calendar year that are documented in the Weeks Report reveals that only 7 involved demands for shorter hours. Five of these were unsuccessful and the results of the remaining two were not known (Weeks 1886, p. 25, Table 4). Demand for shorter hours seem to have played a much more important role in Massachusetts where about 15% of the documented strikes between 1825 and 1879 were to secure shorter hours (Massachusetts Bureau of Statistics of Labor 1880a, pp. 3–71, especially p. 65).[25] The results of these strikes like those elsewhere were not encouraging. Most failed

[24] We refer to this as the "lump of labor hypothesis," that is the proposition that if a firm used say 1000 h of labor per day, this could be supplied by 100 workers working a 10-hour day or by 125 workers working an eight-hour day—a 25% increase in employment as a result of the reduction in hours per day.

[25] See also Pennsylvania Secretary of Internal Affairs (1882, pp. 262–391) and Weeks (1886, pp. 18–19).

and workers either reclaimed their jobs on the same or worse terms or they were replaced (Massachusetts Bureau of Statistics of Labor 1880a; Commons 1918). Rather, most strikes were for higher wages or against a reduction in existing wage rates. Even after 1880 when there seems to have been a growing emphasis upon shorter hours, it still does not appear to have been the dominant issue in most labor disputes.[26]

What direct impact this activity had upon hours is unknown, but hours did decline significantly after 1830. According to estimates assembled by Weeks (1886) and corroborated by independent evidence, workers averaged 11.5 h a day in 1830. By 1850, the average workday had been reduced to 10.9 h. In the course of the next decade, hours declined by as much as in the preceding 20 years to 10.3 h in 1860. During the next two decades they declined much more slowly, averaging fractionally over 10 h a day by 1880.[27] The length of the working day then stabilized during the 1880s before declining slowly after 1890 and very abruptly during the First World War to eight-hours per day (Whaples 1990; Rosenbloom and Sundstrom 1994).

5.4 The Emergence of Protective Legislation

Despite the progressive decline in the length of the working day, direct industrial action seems to have been generally unsuccessful in securing shorter hours even where it had been the primary goal. As a result, labor looked increasingly to political agitation and lobbying for relief. Sometimes the mere threat of legislative action was sufficient to secure a reduction of hours.[28] The widespread acceptance of changes in hours of work necessary for building a political consensus, however, was predicated on new social attitudes and these swung in labor's favor only slowly. It is not difficult to understand why. After all, why should an employer demand less of his workers than the farmer expected of himself or his family (Commons 1918, p. 171–74)? However, the extension of the peak farm-day or farm-week to manufacturing produced a much longer work-year in industry. Moreover, whatever parallel there might have been between agricultural and industrial work disappeared as industry mechanized. In the one, the pace of work was set by man and the rhythms of nature; in the other, by inanimate and tireless machines. In addition, workers were increasingly concentrated in unlighted and poorly ventilated factories. The result was a marked deterioration in the conditions of workers and growing demands for

[26]The increase is complaints about hours is swamped by a dramatic increase in the diversity of workers' demands. In 1880, simple, single issues such as shorter hours, more pay, or against a cut in wages had accounted for almost three-quarters of all labor demands in strikes but by the end of the decade they made up barely half of labor grievances (US Department of Labor 1896, p. 29).

[27]See Atack and Bateman (1992), especially pages 2–22, for the full set of estimates and description of sources.

[28]Sumner (1910, p. 73) argues, for example, that Massachusetts manufacturers reduced hours to prevent enactment of laws.

change from many quarters including social reformers as well as workers directly affected by these conditions. Growing productivity and rising real wages also meant that there was little chance that the mass of workers and their dependents would fall below subsistence and so these new arguments were free to come to the fore (Towles 1908, p. 7).[29]

However, it may also be significant that much of the agitation for shorter hours and limits upon employment of women and children only began when the bargaining power of those directly affected had weakened. Adult male workers were thus in a much stronger position to press their case for shorter hours without competition from women and children and employers had less incentive to resist. Second, support for protective legislation for women and children often came from adult male-dominated labor organizations. For example, the New England Working Men's Association meeting at Lynn in January 1846 offered a resolution "in behalf of the Lowell factory girls" for adoption of the ten-hour day (Commons 1918, p. 540). Commons viewed this as fraternal cooperation between labor organizations on behalf of those not yet organized but it is equally consistent with the rent-seeking hypothesis.

The feasibility of a political solution was clearly demonstrated when President Van Buren issued an Executive Order in 1840 establishing a ten-hour day—without reduction in pay—for all federal employees.[30] The lesson was further reinforced by the adoption of the eight-hour day for federal workers in 1869.[31] Rather than securing shorter hours through bargaining or industrial action on a workplace-by-workplace basis, legislation secured through political action represented a comprehensive and effective solution. The emerging coalition of labor with social, civic, philanthropic, and church groups; social reformers and humanitarians; and legislators and bureaucrats also made legislative action more certain and less risky for the individual worker. This coalition was to prove effective in state after state in securing legislation limiting hours, particularly after Massachusetts had shown the way.

The first state law setting hours of work was passed by New Hampshire in 1847. It established 10 h as a legal day's work for everyone and provided that no person could be required to work longer "except in pursuance of an express contract requiring a greater time."[32] No provisions were made for enforcement. Nor were there any specific penalties for violators. Nevertheless, just 3 days before the law was to go into effect, manufacturers submitted the necessary express contracts to their employees. All who refused to sign—including between one half and two-

[29] For evidence on rising productivity and real wage growth see Sokoloff (1986, pp 679–729, especially Tables 13.4–13.6).

[30] Executive Order of March 31, 1840 (Richardson 1908, p. 602). For an extended discussion of the misquotes of this Executive Order by a wide variety of labor historians from Richard Ely to John R. Commons see Kelly (1950).

[31] See the discussion surrounding this in Cahill (1968, p. 69–71).

[32] New Hampshire, 1847 Session Laws, Ch. 4.

thirds of the operatives at the Nashua Corporation—were fired and replaced.[33] One might therefore argue that such a law was totally ineffective. Similar laws were also passed in other states—Pennsylvania (1848), Maine (1848), New Jersey (1851), New York, (1853), Rhode Island (1853), California (1853), Georgia (1853), and Connecticut (1855).[34] Many improved upon the New Hampshire law by specifying penalties for violators. However, all preserved the provision that not only guaranteed the primacy of existing contracts but also imposed no limitations upon future agreements.

Such flawed legislation could, however, be made effective through collective action. Following passage of 10-hours legislation by the Pennsylvania legislature in March 1848, seven cotton mills in Allegheny shut down on July 4, 1848, after employees demanded a reduction in hours to ten. The strike/lock-out was eventually settled on August 28 with workers winning the 10-hour day but with a 16% reduction in daily wages.[35] Later in the year, however, at least one mill rescinded the pay cut (Pennsylvania Secretary of Internal Affairs 1882). Similar action also occurred in parts of New Jersey.[36] Nor did these strikes settle the issue one way or the other. For example, in 1866 all the large cotton mills in Allegheny City, Pennsylvania were again running 11.5 h a day despite 10-hour strikes earlier in the year (Sumner 1910, p. 72).

It is claimed that the loophole provided by the contracting-out provision of the declaratory laws establishing 8 or 10 h as a legal day rendered them ineffective (Commons 1918). This would seem doubly so since the Courts assumed the existence of a contract superseding the law whenever customary hours were longer.[37] One might therefore reasonably ask, why would legislatures pass such patently flawed laws? One possibility is that they sought a sop to labor and the coalition of social reformers without antagonizing manufacturers. Another explanation is that the laws were not made effective because they simply codified and ratified the existing situation. Some large mills in Lowell and Lawrence, for example, had adopted the 10-hour day in the late 1860s, years before the passage of Massachusetts' ten-hour law (Sumner 1910, p. 72). However, this was certainly not always the case. Women in most Massachusetts mills were still employed 11 h

[33] Sumner (1910, p. 69) quoting *Voice of Industry*, 3 September 1847 and *Voice of Industry*, 17 September 1847. See also Commons and Andrews (1927, p. 248).

[34] See California: California, Session Laws of 1853, Ch. 131, p. 187; Connecticut: Connecticut, Session Laws of 1855, Ch. 45; Georgia: Georgia, Code of 1861, Sec. 1847; Maine: Maine, Session Laws 1848, ch. 83; New Jersey: New Jersey, Session Laws 1851, pp. 321–322 (This law only applied to cotton, woolen, silk, paper, glass, and flax factories and to iron and brass works); New York: New York, Session Laws 1853, ch. 641; Pennsylvania: Pennsylvania, Session Laws 1848, Act 227; Rhode Island: Rhode Island, Session Laws 1853, p. 245.

[35] Pennsylvania Secretary of Internal Affairs (1882, pp. 272–3). Since we do not know how many hours a day were worked before the lockout, we do not know whether or not the reduction changed the hourly wage rate.

[36] See Sumner (1910, p. 70) and Commons and Andrews (1927, p. 248).

[37] *HELPHENSTEINE V. HARTIG* (5 Ind. App. Ct. 172, 1892).

a day until prevented by legislation (Sumner 1910, p. 73). Similarly, workers in New Hampshire flour mills averaged about 10.4 h a day in 1880—more than 30 years after the passage of that state's ten-hour law.[38] However, there was also a very sound legal reason why the laws were framed as they were: without the loophole these laws were viewed as an unconstitutional interference with the right to free contracting. Nor was this legal question regarding constitutionality unreasonable: In 1905, the U.S. Supreme Court in *LOCHNER v. NEW YORK* stuck down a New York state law setting maximum hours for bakers.[39] The law was held invalid as an unreasonable interference with the right of free contract and an excessive use of the police powers of the state. The case is also notable for the dissent of Justice Holmes who argued that the Constitution was "not intended to embody a particular economic theory, whether of paternalism . . . or of laissez-faire."[40]

However, we believe it would be a mistake to assert that these laws had no effect upon hours because of this loophole. To many, the law represents a moral force, we would therefore expect that some employers at the margin might have been moved to shorten hours.[41] Compliance is not always based upon the risk of detection and the threat of punishment. Nevertheless, these laws must not have achieved all their goals because states continued to seek a better way to mandate hours for women and children.

Except for the Georgia law that applied only to white labor, the laws establishing the length of a "legal day" made no distinctions between workers. Either all were protected or, as we have seen, none were. The passage of more discriminating laws, however, was to prove easier and the laws themselves may have been somewhat more successful in achieving their stated goal. In particular, it was also argued that women and children needed protection and deserved preferential treatment because the health of future generations depended upon them. Moreover, in the case of women, while "ordinarily men can rest when their day's toil is over, . . . there are few working girls who do not have at least mending and laundering to do in the evenings, and married women must take the entire care of their homes and children before and after work" (Commons and Andrews 1927, p. 202). Willingness to express such paternalistic concern, however, may simply reflect the declining relative importance of women and children in manufacturing industry (Goldin and Sokoloff 1982). For the men, on the other hand, shorter hours where deemed desirable were thought of in terms of increased leisure, although emphasis was sometimes given to family life.[42] More often, shorter hours for men seem to have been thought undesirable, simply

[38] Unpublished sample data from Atack and Bateman (2004b).

[39] 198 US 45.

[40] 198 US 45.

[41] For those readers inclined to question this assertion, we ask that you reflect upon the actions of those who, finding lost property—sometimes large sums of untraceable cash—return it, or those who follow posted speed limits.

[42] See, for example, Ohio Bureau of Labor Statistics (1879, pp. 280–86), where a carriage manufacturer reported: "I am strongly in favor of an eight-hour system. Workingmen would be better informed if they had more rest, and would be more healthy and longer lived . . . The man

providing more drinking time (Ohio Bureau of Labor Statistics 1879). Universal shorter hours for men thus had to wait until well into the twentieth century (and Prohibition), while women and children received protection much earlier.

Instead, men decided to "fight the battle from behind the women's petticoats" (Commons 1918, p. 462). In 1852, Ohio became the first state in the nation to enact protective legislation specifically limiting the hours of work for women. The law provided that women could not be compelled to work longer than 10 h a day under penalty of a $5–50 fine.[43] It became the model for similar legislation passed by Minnesota, Wisconsin, and Dakota.[44] The stipulation that women could not be compelled to work longer hours, however, supposedly rendered these laws ineffective regardless of enforcement and penalties. Commons and Andrews argue that "as most employees will voluntarily work for twelve or more hours a day when they cannot find any one to employ them for 10 h, the law became almost entirely inoperative" (Commons and Andrews 1927, p. 249). Alternatively, one might argue that the exception permitted those whose tastes were different to work longer hours while those who did not want to could not be forced to work more than 10 h. Under this interpretation the exception was Pareto efficient though in practical terms it generally required firms to make different hours offerings and so was no different from the situation that should have existed in the absence of regulation.

The main battleground in the fight for shorter hours, however, was Massachusetts where manufacturers had long played the lead role in determining the hours of work:

> In general, the hours of labor in Massachusetts, in spite of the lack of legislation, were reduced first, other States following. When the mills of Massachusetts ran 12 hours a day, those of Rhode Island and New Hampshire ran 13 hours. When her mills came down to 11 hours a day, theirs came down to 12. (Sumner 1910, pp. 72–3)

Petitions requesting passage of a ten-hour law in Massachusetts had been presented to the state legislature in 1842, 1843, and 1844 but nothing came of them (Kingsbury 1911, pp. 1–129, especially pp. 24–27). Nevertheless, agitation and lobbying continued, bearing some fruit in the adoption of the 10-hour day by a few mills in the 1860s although the norm remained 11 h or longer (Sumner 1910, p. 72).

Manufacturers subsequently further reduced hours in a vain effort to forestall legislation (Sumner 1910, p. 73). Despite this last ditch effort the Massachusetts legislature passed a law in 1874 that circumvented the traditional problems by arguing that, since the health of future generations depended upon the health of the mothers, the regulation of hours of work for women fell within their police powers—a view with which the courts concurred. This law therefore superseded

---

who works ten hours, and does his home chores . . . is too tired to study or in any way tax his mind." (p. 281).

[43] Ohio, Session Laws 1852, v. 50, p. 187.

[44] Dakota: Dakota Territory Legislature. Dakota Session Laws 1862-3, ch. 49; Minnesota: Minnesota, Session Laws 1858, ch. 66; Wisconsin: Wisconsin, Session Laws 1867, ch. 83.

Table 5.3 State maximum hours laws affecting women in 1879–1880

State	Year passed	Per day	Max hours per week	Enforcement	Contracting out allowed?
California*[a]	1853	10	-	None	Yes
Connecticut*[a]	1855	10	-	None	Yes
Dakota	1863	10	-	$10–100	Yes
Florida*[a]	1874	10	-	None	Yes
Georgia[b]	1853	Daylight	-	$100	Yes
Illinois*[a]	1867	8	-	None	Yes
Maine*[a]	1848	10	-	$0–100	Yes
Massachusetts	1874	10	60	$0–50	**No**
Minnesota	1858	10	-	$10–100	Yes
Missouri*[a]	1867	8	-	None	Yes
New Hampshire*	1847	10	-	None	Yes
New Jersey*[a]	1851	10	-	None	Yes
New York*[a]	1853	10	-	None	Yes
Ohio[c]	1852	10	-	$5–50	Yes
Pennsylvania*	1848	10	60	None	Yes
Rhode Island*	1853	10	-	$20	Yes
Wisconsin	1867	8	-	$5–50	Yes

Notes: * indicates that the law made no distinction between men and women
[a]The law established the length of the "legal day"
[b]The law applied to white labor only
[c]The law was repealed in 1879 effective January 1, 1880. For sources see text

contracts.[45] It also contained provisions for penalties against employers who were found guilty of "willfully employing" women for longer hours. Although some have argued that this stipulation weakened the law because willful intent was difficult to prove, at least one offender was successfully prosecuted and convicted under the statute and that conviction survived appeal. Nevertheless, in 1879 the law was rewritten to strike the word "willful," making prosecution easier.[46] Thus, by 1880, women were covered directly and indirectly by a wide variety of different laws regulating their hours of work. These are summarized in Table 5.3.[47]

[45]Massachusetts, Laws 1874, C. 74. The same principle was eventually established at the federal level in 1908 in *MULLER v. OREGON* (208 US 412).

[46]See, for example, *Report of the Convention of the International Association of Factory Inspectors*, 1894, p. 65 quoted in Goldmark (1912, p. 213).

[47]There is, apparently, no complete summary of legislation regulating hours for work. Table 5.3 has been compiled from a wide variety of sources, among them US Women's Bureau (1931), Commons (1918, p. 541–5), and Sufrin and Sedgwick (1954, p. 78). US Women's Bureau (1931), for example, misses the early legislation in California, Connecticut, Georgia, New York, and Rhode Island. Also Persons (1911), Baker (1925, p. 109), Towles (1908), and Edwards (1907).

Laws regulating children's hours presented none of the legal dilemmas of those protecting women. As minors, children could not legally enter into contracts. Moreover, the principle that minors could be treated as wards of the state was enshrined in the English common law tradition. As a result, the constitutionality of protective legislation covering children was not questioned, and the legal obstacles that stood in the way of legislation protecting men and women did not exist. The appeal to humanitarian concerns was also much easier for children. For example, the 1832 convention of the New England Association passed a resolution declaring, in part, that:

> Children should not be allowed to labor in the factories from morning till night, without any time for healthy recreation and mental culture; as it not only endangers their own well-being and health, but ensures to the country the existence of a population, in the approaching generation, unfitted to enjoy the privileges and to exercise the duties of citizens and freemen (Commons 1918, p. 321).

Others stressed the evils of the workplace:

> Worse than physical hardship, more blighting than cold or ill-treatment is the inevitable insight of the childish mind into duplicity and vice. A gradual hardening of the sensibilities ensures from constantly hearing words unfit for the ears of youth witnessing the degrading acts and ugly passions which are too frequent in some work rooms where the sexes indiscriminately mingle. Little beings that should be sheltered by mothers' love are only taught the alphabet of sin (De Graffenried 1890, p. 198–9).

Consequently, laws protecting children were passed by many states quite early on—earlier in fact than other protective labor legislation. Connecticut and Massachusetts passed laws in 1842 limiting children to no more than 10-hours work per day.[48] By 1879, thirteen states had such laws on their books.[49]

Legislatures constrained—"protected"—juveniles in other ways, too. Most common was the passage of compulsory education laws. According to Landes and Solmon (1972), by 1879, 16 state had enacted laws that typically required at least 3 months school attendance a year for children under 15, with not less than 6 weeks of the attendance in consecutive weeks (Table 5.4).[50] These laws complicate our analysis. Rhode Island, for example, passed a law in 1840 requiring that children under 12 (revised to 15 in 1854) attend school for at least 3 of the 12 months preceding employment but not until 1853 did the state legislature pass a law prohibiting the employment of children under 12 and limiting the hours of work of those aged 12–15 to a maximum of 11 h a day (Towles 1908, p. 25). They thus

[48]Connecticut: Connecticut, Laws 1842, C. 28; Massachusetts: Massachusetts, Laws 1842, C. 60.

[49]This list is assembled from Ogburn (1912), specifically from Table 30, pp. 108–9 but corrected for the omission of Maine that had passed compulsory schooling legislation in 1875. See also Table 1, Landes and Solmon (1972) quoting Ogburn (1912). The list in Landes and Solmon (1972), however, includes Indiana, Iowa, and Colorado, which do not appear in Ogburn's compilation, but excludes Connecticut and Illinois.

[50]The states were California, Connecticut, Kansas, Maine, Massachusetts, Michigan, Nevada, New Hampshire, New Jersey, New York, Ohio, Vermont, Washington, Wisconsin, and Wyoming. See Landes and Solmon (1972, p. 56–7).

Table 5.4 State compulsory education and child labor laws in 1879–1880

	Education law			Child labor law		
State	Year	Ages	Requirement	Year	Max. hours	Ages
California	1974	8–14	16 weeks	–	–	–
Connecticut	1872	8–14	in sess.	1857	10/58	<15
Indiana	–	–	–	1867	10	<16
Kansas	1874	8–14	12 weeks	–	–	–
Maine	1875	8–15	16 weeks	1848	10	<18
Maryland	–	–	–	1876	10	<16
Massachusetts	1852	8–14	20 weeks	1842	10/60	<18
Michigan	1871	8–14	4 months	–	–	–
New Hampshire	1871	6–15	12 weeks	1846	10	<14
New Jersey	1875	7–16	20 weeks	1852	10	<14
New York	1874	8–14	14 weeks	–	–	–
Ohio	1877	8–16	20 weeks	1852	10	<14
Pennsylvania	–	–	–	1848	10/60	<21
Rhode Island	1854	7–15	3 months	1853	11	<15
Vermont	1867	8–14	20 weeks	1868	10	<15
Washington	1871	8–18	3 months	–	–	–
Wisconsin	1879	7–15	12 weeks	1853	10	<14

Note: For sources see text

denied some children the opportunity of work in factories, but the laws did not apply to all minors. As a result, compulsory education laws—to the extent that they were effective—probably affected the distribution of the labor force between adults and children more than hours legislation.

5.5 Some Qualitative Evidence on Enforcement and the Impact of Hours Laws on Behavior

We do not know how effective many of the laws were, although we do know that workers in parts of Pennsylvania and New Jersey successfully resisted efforts to force them to sign explicit contracts for long hours of work in advance of laws setting a shorter maximum. On the other hand, hundreds of employees in New Hampshire were discharged and blacklisted over their refusal to sign similar contracts in advance of the New Hampshire 10-hour law.[51]

We have suggested above that the law represents a moral force. Some people change their behavior to comply with legal requirements regardless of whether

[51]These blacklists were also circulated throughout New England, making it difficult for these workers to find new work (Sumner 1910, pp. 69–70, 94–6).

or not the law is enforced and what penalties might apply. There exist citizens whom we describe as "law-abiding" not "law-fearing." We are, however, reluctant to rest our case upon this argument, especially in light of the repeated reference to provisions for enforcement, penalties and the existence of loopholes.

Much of what we know about the impact of hours, child labor and compulsory education laws, unfortunately, comes from just for one state—Massachusetts. The evidence, though circumstantial and qualitative, suggests that these laws were effective in reducing the hours of women and limiting the employment of children and that these laws were enforced. It is also, as we will show, consistent with our quantitative evidence.

The circumstantial evidence on hours of work, assembled from Weeks (1886), is summarized in Fig. 5.1.[52] This shows the average hours of work in Massachusetts textile mills, a Massachusetts paper mill and all manufacturing in that state at 5-year intervals between 1830 and 1880 together with the average hours in all U.S. industry. We would have liked to show the hours in boots and shoes (another important employer of women) as well but no Massachusetts boot and shoe manufacturer gave a complete history of their hours of work to Weeks. Until 1875, hours of work in the textile and paper mills were, on average, much longer than hours in other industries. Thereafter, they were shorter. This reversal coincides with passage of Massachusetts' 10-hour law in 1874. The textile industry, while not the only employer of women in the state, was undoubtedly the one most directly and seriously affected by the passage of this law. In 1870, the industry employed more than 30,000 women—more than a third of all female employees in the state—and perhaps as much as 62% of its employees were covered by the law (US Census Office 1872, p. 430 and 489, Table 8). Another 2300 women were employed in the Massachusetts paper mills in the Connecticut River valley (US Census Office 1872, p. 463).

Our belief that dramatic change in hours in Massachusetts textile mills after 1874 was a direct result of the passage of the 10-hour law is further supported by contemporary statements of manufacturers. For example:

> From 1830 to 1865, all classes worked 13 hours a day; from 1865 to 1875, 11 hours, and since 1875, 10 hours. *The reduction to 10 hours was made because the state allows women and minors to work but 60 hours a week and the proprietor desired all hands work the same time.* (Emphasis added) (Weeks 1886, p. 394)

The decision to reduce hours for everyone is consistent with employer monitoring to avoid shirking that included the regulation of entry and egress from the factory[53] and the establishment of rigid work rules including hours (Illinois Bureau of Labor Statistics 1886, pp. 501–26, especially 506).

[52] Weeks Report, op. cit. Weeks (1886).

[53] David Montgomery quotes the Senate testimony of R. D. Layton that the five hundred workers at the carriage works of James Cunningham and Sons in Rochester, New York, were locked in from starting time to quitting time so that no individual pieceworker could finish a stint and leave. See Montgomery (1989, p. 152).

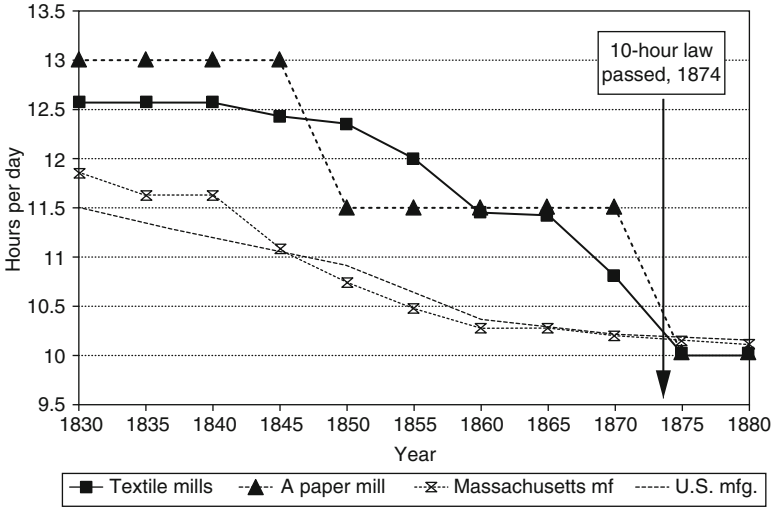

Fig. 5.1 Hours of work in Massachusetts industry

Our assertions of the existence of "law-abiding" citizens notwithstanding, compliance was not automatic. As we have noted, the Hamilton Manufacturing Company was successfully prosecuted for violating the law. Moreover, some businesses continued to violate the law even after this successful test of the statute. For example, during 1886, in 377 factory inspections, 27 violations were found, four of them deemed worthy of prosecution.[54] In his report, the Chief of the District Police remarked:

> In former reports I had occasion to mention the attempt that had been made to evade the [1874] statute and to secure an extension of the daily hours in certain factories, upon the pretext that the time so taken was required to get up a proper rate of speed of machinery preparatory to the process of production. The margin over ten hours thus claimed and constantly taken by mill superintendents and agents *in some sections of the Commonwealth* amounted to twenty minutes or more daily. The difficulty of obtaining evidence was not lessened by the natural desire of the operatives to retain their employment by withholding information which might incriminate employers. *In every case where there appeared sufficient evidence to show that there had been a substantial violation of the law and the warning of our officers had not been heeded, I felt it my duty to make complaints in court* ….(emphasis added) (Massachusetts District Police 1887, p. 18)

The inspector for Middlesex County (which includes Boston and Lowell and home to the Hamilton Manufacturing Company) reported some continuing problems:

[54]Massachusetts District Police (1887, p. 12). Most of the violations were apparently minor technical violations of the section requiring posting of a notice stating the hours of work that was amended during 1886. See Massachusetts, Laws 1886, C. 90.

> Nearly all the mills are now running by a schedule arranged for 60 hours a week, several having changed their time from 65 or 66 hours to 60 within a year or two.
> Occasional violations are detected, or reported to me, where there is temporary overwork to fill orders at the stipulated time, or to catch up in some department. In some cases, where the engine is started several minutes before the schedule (sic) time, complaints are made that the women, generally weavers, go to work five or ten minutes before the proper time, thus gaining one or two hours a week.
> Violations of this latter class are much more difficult to deal with than the other. The women want to work, and usually remember that they lose sufficient time during the day to make up for the time gained morning or noon. (Massachusetts District Police 1887, p. 55)

In general, however, by 1886, the inspectors thought that there was general compliance with the law. Consider, for example, the following remarks: from Worcester County: "I find little cause for complaint; the manufacturers in my district are generally strictly complying with the law (Massachusetts District Police 1887, p. 63);" from Hampden and Hampshire counties: "more fully complied with during the last 2 years than ever before (Massachusetts District Police 1887, p. 71);" from Bristol, Barnstable, Dukes and Nantucket counties: "The law …is much better observed than formerly (Massachusetts District Police 1887, p. 69)." Lastly, no problems whatsoever were reported for Essex County which encompasses the textile center of Lawrence and the shoe-making cities of Lynn and Danvers.

The laws regulating the employment of children and requiring school attendance almost invariably contained enforcement and penalty provisions. They also seem to have been frequently violated, perhaps because prosecution was uncertain even when violators were identified. For example, in his third report as Massachusetts Commissioner of Labor, Carroll Wright drew attention to the situation at Pacific Mills at Lawrence for which:

> no children are returned, although it is *known by investigation that children under 15 and many under 10 years of age are constantly employed in this establishment, contrary to law*. This is a well-known fact in Lawrence, and an effort was made through the State Police to remedy this evil, but without effect. …*Our investigations warrant us saying that the School Law is universally broken.* (Massachusetts Bureau of Statistics of Labor 1872, pp. 162–3) (emphasis added)

At the time, Pacific Mills was the largest manufacturing enterprise in America. Fifteen years later, the inspector for Middlesex County wrote:

> In regard to the law regarding the employment of minors in factories, I think it can be safely said that few laws on the statute book are better complied with …
> When I began work in this department, seven years ago, not one in ten of the children under fourteen years of age employed in factories had certificates on file to show they had attended school as required by law …
> Now, ninety-nine out of a hundred such children have properly signed certificates showing the required attendance. Then, it was a common thing to find children under ten years of age at work in the mills; now, none under twelve are employed. (Massachusetts District Police 1887, p. 55)

Others echoed similar sentiments: "during the past year I have had no occasion to enter prosecution for the employment of children under 16 years of age. In fact, it is rare to find children under fourteen employed" (Massachusetts District Police 1887, p. 63).

Unfortunately, we have found very little information as yet about what happened outside of Massachusetts. What evidence we have found suggests that, while little or no effort was made to enforce hours laws for adults outside of Massachusetts, some effort was made to ensure compliance with the child labor and child education laws.

5.6 Recent Debate Over Female Protective Legislation

In a 1980 study, Landes (1980) concluded that the legislation protecting women by limiting hours of work had two effects: First, it dramatically shortened work hours for women, but second, it also limited employment opportunities for women. Those conclusions have been questioned by Goldin (1988). She argues that Landes' model is mis-specified in that it forces the entire impact of restrictions in women's hours to fall on women hours and denies the possibility that men might have gained from the restrictions. By removing implicit restrictions from the equation, Goldin shows that protective legislation reduced the hours of all workers and that employment losses were not dependent upon the restrictiveness of the legislation.

Aside from the question of misspecification, four serious questions can be raised regarding these important studies:

- Both Landes and Goldin rely upon aggregate state data. As a result, they cannot take account of the variations in hours and gender mix from industry to industry.[55] For example, whereas women comprised more than half of the labor force in textiles and clothing, they typically made up less than 1% of the workers in the construction trades and primary metals (US Census Office 1928). Moreover, there were small but statistically significant hours variations from industry to industry (Atack and Bateman 1992).
- The model used by Landes gives equal weight to each state so that the law on women's hours in New York which covered 351,450 women in 1920 and may have affected 876,680 men receive equal billing with those in, say, North Dakota where they only affected 4472 workers (Atack and Bateman 1992, p. 37–8).[56]
- Both Landes and Goldin regress labor hours and employment share of women in 1920 on a legislative dummy variable that measures whether or not a state had enacted protective legislation prior to 1914. Contemporaneous legislative data were not used because by 1920 virtually all states had enacted protective legislation (Landes 1980, p. 483–83). This dichotomy in the data is doubly

[55]Manuscript census data for manufactures later than 1880 are not available. The 1890 data were destroyed by a fire at the Department of Commerce in 1921 and later data, to the extent that they might have survived, have been subject to privacy rulings by the Census Bureau and the National Archives. This situation may be changing with the recent access granted Timothy Bresnahan and Daniel Raff to the manuscripts of the 1929 and 1931 censuses.

[56]The 1920 Census lists the 1919 employment of women in North Dakota as only 336.

questionable in light of the dramatic decline in hours that took place between 1914 and 1920. In 1914, female production workers averaged 50.1 h per week; in 1920, they averaged 43.0 h. For men, the change was much less dramatic, average weekly hours only declined from 51.5 h to 48.2 h (US Department of Commerce 1975, Series D 834 and D 837).

- Both authors use a single dummy variable to capture the diverse array of legislative efforts to protect women and measure their impact upon hours. This implicitly assumes that all laws were created equal. They were not in 1914. Nor were they in 1920. In 1914, the limits on hours of work for women varied from between 8 and 10.25 h a day and between 48 and 70 h a week.[57]

5.7 Protective Legislation and Hours of Work

Our model elaborates upon that used by Landes and Goldin. Consider the following identity for the firm:

$$H \equiv \alpha_f H_f + \alpha_y H_y + (1 - \alpha_f - \alpha_y) H_m \tag{5.1}$$

where $0 \leq \alpha_f, \alpha_y \leq 1$, and $\alpha_f + \alpha_y \leq 1$. This simply expresses the average daily hours of work for the firm as the weighted average of male hours (H_m), female hours (H_f) and child hours (H_y), where α_f and α_y are the proportions of the firm's labor force that was female and child. This may be rewritten as:

$$H = H_m + \alpha_f (H_f - H_m) + \alpha_y (H_y - H_m) \tag{5.2}$$

which now expresses hours of work in terms of male hours of work and the weighted differences between male and female or child hours.

Suppose now that a law, represented here by LAW, is passed restricting the hours that one group, say women, can work. Let the maximum hours of work that women are permitted to work under this law be H_{fmax}. If $H_{fmax} > H_f$ then the law is irrelevant for the firm and has no effect upon it. If, however, $H_{fmax} < H_f$ then the firm is in violation of the law and must decide whether to risk prosecution or comply. Whether or not the firm chooses to risk prosecution presumably depends, in part, upon the probability of detection and conviction and the size of the likely

[57] See Landes (1980, pp. 482–3), especially Table 1. Note, however, that this table is incomplete and inaccurate because of errors in US Women's Bureau (1931). For example, the Massachusetts law protecting women was first passed in 1874 (not 1879 as shown in Table 1 of Landes 1980). It was amended in 1879, dropping the word "willful", to make prosecution of violators easier. Moreover, a large number of states such as California, Connecticut, Georgia, Maine, New Hampshire, New York, Ohio, Pennsylvania, and Rhode Island had passed laws regulating hours at earlier dates than those shown. See Table 5.3.

penalty relative to the gains versus the costs of complying. This will be determined by factors such as:

- whether or not there is any mechanism for enforcement, such as factory inspectors who had the right to make unannounced visits, review records and interview employees;
- the risk that this firm would be visited by inspectors—Was it, for example, in an industry known traditionally to have employed the affected group for longer hours than now permitted by law? Did the firm draw attention to itself by vocal opposition to the law?;
- the possible seriousness of the violation measured, for example, by how many people were affected; the visibility of the firm—its size, location, etc.;
- the size of the maximum penalty upon conviction;
- how easily adjustments could be made through, for example, the substitution of other factors of production to compensate, the rearrangement of schedules, or changes in employment either to eliminate the affected factor or even increase the numbers employed so as to maintain the same total number of labor hours for this factor.

This list is not intended to be exhaustive. Suppose that the law is effective and the firm complies by reducing the hours of work for women covered by the law from H_f to H_{fmax}, or by $H_{fmax} - H_f$. As a result, daily hours of work for the firm will be reduced by $(H_{fmax} - H_f)\alpha_f$ hours and the marginal impact of this law on women's hours, q, may be approximated:

$$q = (H_f - H_{fmax})/H_f \tag{5.3}$$

We can now rewrite Eq. (5.1) to include firms operating in states with and without laws setting maximum hours of work for women by including the dummy variable LAW that assumes a value of 1 if the state has a maximum hours law for women and zero otherwise:

$$H = \alpha_f H_f + \alpha_y H_y + (1 - \alpha_f - \alpha_y) H_m + q H_f \alpha_f (LAW) \tag{5.4}$$

The specification, however, assumes that because the law applies only to women, it only affects women hours. This contradicts the assertions of contemporaries. Consider again the following two quotes:

> Ten hours' labor per day has been the recognized work-day in nearly every mechanical and manufacturing industry in the United States, for the past thirty or forty years, except in the textile manufactories of the eastern States; and in them, as in England, *legislation has been invoked specially for women and children, and which, being secured, caused a lessening of the hours of adult male labor* ….(Ohio Bureau of Labor Statistics 1879, p. 266, emphasis added)

> From 1830 to 1865, all classes worked 13 hours a day; from 1865 to 1875, 11 hours, and since 1875, 10 hours. *The reduction to 10 hours was made because the state allows women and minors to work but 60 hours a week and the proprietor desired all hands work the same time*. (Weeks 1886, p. 394)

An obvious question is why "the proprietor desired all hands work the same time." We believe that this reflected complementarity between men, women and children in nineteenth century factories, resulting from rigid gender- and age-based job discrimination. There was "women's work" and there was "men's work," and only rarely did the 'twain meet.

In a study of the boot and shoe industry, for example, it was asked: "Why should cutting continue to be so exclusively man's work?" Their answer:

> Long continued custom is no doubt a weighty reason. ... Yet custom is not all. The material handled in the cutting room is heavy and clumsy, while the manipulation of hammer or machine requires considerable muscular force, which the operator must exercise while standing. Women brought up in towns will shun any operation to which these conditions are attached. There are few exceptions ... in the lasting room, there is no question of its unfitness for women .... (US Department of Labor 1915, p. 34 and 39)

The result was:

> men and women do not often work in competition in the same industry. Occupations are apt to be assigned to one sex or the other, and even when both work nominally at the same occupation there is apt to be a difference in the kind of work done or the methods employed. (US Department of Labor 1915, p. 40)

Such historical statements can be found in industry after industry: In canning and preserving: "there was a well-defined line of demarcation between the occupations of men and women ...;" in cans and boxes: "Competition between the sexes does not seem to have arisen to any serious extent ...;" in confectionery: "Candy making is sometimes called a woman's industry, because although men do the actual making, the women, who do everything else, outnumber them so greatly ...;" in core making: "since women are used only for making the lighter and less difficult cores, and since they never learn the trade as a whole, there seems little prospect that they will ever prove dangerous rivals of men in this work ...;" in paper box making: "in the highly organized factories the men and women had different occupations, so that competition did not exist ...;" et cetera, et cetera (US Congress, Senate 1910, p. 42, 58, 129, 140, and 243).

If, then, men and women had different occupations and played different role within the firm, they were complementary factors. In highly-integrated operations, this interdependence would dictate common starting and stopping hours. So too would ease of administration, supervision and the maintenance of discipline and control. Consequently, shorter hours for women very probably also translated into shorter hours for men—an outcome for which men had often agitated, lobbied, and struck.

This so, we can capture the impact in our model by approximating the marginal impact of hours legislation for women upon men by:

$$r = (H_m - H_{fmax})/H_m \tag{5.5}$$

and the marginal impact of hours legislation for women upon children by:

$$s = (H_y - H_{fmax})/H_y \tag{5.6}$$

The impact of these changes on the firm's hours is the given by:

$$r H_m(1 - \alpha_f - \alpha_y) \tag{5.7}$$

and

$$s H_y \alpha_y \tag{5.8}$$

We can now rewrite Eq. (5.4) to admit the possibility that adoption of an effective law governing women's hours may affect the hours of men and children as well as those of women:

$$H = \alpha_f H_f + \alpha_y H_y + (1 - \alpha_f - \alpha_y) H_m + q H_f \alpha_f (LAW)$$

$$+ s H_y \alpha_y (LAW) + r H_m (1 - \alpha_f - \alpha_y)(LAW) \tag{5.9}$$

where the first three terms are the contribution of female, youth and male hours to the firm's hours and the last three terms represent the impact of an hours law for women upon each of these contributions.

Rearranging terms in Eq. (5.9):

$$H = H_m \alpha_f (H_f - H_m) + \alpha_y (H_y - H_m) + r H_m (LAW)$$

$$+ \alpha_f (q H_f - r H_m)(LAW) + \alpha_y (s H_y - r H_m)(LAW) \tag{5.10}$$

expresses the firm's average daily hours in terms of the hours worked by men each day adjusted by the weighted difference between female or child and male daily hours, the impact of a maximum hours law for women upon male hours per day and the weighted difference between female or child and male hours per day under a maximum hours law for women. This impact will vary not only with terms of the law but also its effectiveness. If the law is ignored, its impact will be zero.

Male, female and child hours, however, were not reported by the Census. As a result we cannot estimate H_m, H_f, and H_y directly or examine the difference female or child hours per day and those for men. Instead, the census simply reported daily scheduled hours for the firm as well as the average number of men, women and children employed. These yield estimates of H, α_f and α_y. Though unobserved, we can estimate the scheduled hours of work for separate groups and determine whether there is any systematic association between these hours and the passage of maximum hours laws. Consider the OLS regression equation based upon Eq. (5.10) above of the form:

$$H = \beta_0 + \beta_1 \alpha_f + \beta_2 \alpha_y + \beta_3 (LAW) + \beta_4 (\alpha_f LAW) + \beta_f (\alpha_y LAW) + u \tag{5.11}$$

The constant term, β_0, is an estimate of the average daily hours for men in states with no maximum hours law; β_3 is an estimate of the impact of a maximum hours law for women on the average daily hours for men; and β_1, β_2, β_4, and β_5 are estimates of the difference between female or child daily hours and male daily hours in states without and with a maximum hours law for women. A statistically significant estimate for β_3 would then be evidence in support of the Ohio Commissioner of Labor's view that female hours legislation was associated with a shorter workday for men. Statistically insignificant estimates for β_1, β_2, β_4, and β_5 would be evidence that employers required men, women and children to work the same hours. A statistically significant negative estimate for β_4 would imply that, in those states where there was a maximum hours law for women, women worked shorter hours than men. Nowadays this would be referred to as a difference-in-differences estimator. It is important to note that we make no inferences regarding cause and effect based upon the regression results. We do not interpret the results as showing that protective legislation limiting the length of the working day for women led or did not lead to shorter hours for men, women or children; merely that shorter hours were, or were not, associated with such legislation. It is the contemporary evidence we have uncovered that makes this linkage explicit.

5.8 Impact of Hours Legislation in 1920 Compared with 1880

Ignoring for the moment the question of child labor and the impact of child labor and compulsory education laws upon the labor force composition and hours, Table 5.5 reproduces Goldin's replication of Landes' equation and the results from her respecified model.

In Goldin's model, scheduled hours were regressed on the percent of the industrial labor force that was female, the interaction between this variable and a dummy representing whether or not a state had adopted a maximum hours law for women, the maximum hours law dummy variable, the percent of the state population that was urban and a dummy variable for the southern states. In Landes' specification, the maximum hours law dummy variable was not included. These estimates for 1920 were made using state-level data. The same variables with firm-level data were used in the 1880 equations except that urban was a dummy variable for whether or not the firm was located in an urban area, defined by the Census as an incorporated town or city with at least 2500 inhabitants.

Landes concluded that women worked longer hours than men $\beta_1 > 0$ but that women's hours were sharply reduced in those states that had adopted a maximum hours law for women $\beta_4 < 0$. Goldin's respecification, however, suggests that male and female hours were the same ($\beta_1 = 0$ and $\beta_4 = 0$) but that hours for both men and women were substantially lower in states that had passed a maximum hours law

Table 5.5 The impact of hours legislation for women on scheduled weekly hours in 1920 compared with 1880

Independent variables	1920 Landes model	1920 Goldin model	1880 Landes model	1880 Goldin model
Constant (β_0)	53.4	54.8	63.6	63.6
	(80.7)	(88.1)	(536.4)	(508.0)
South	1.34	1.51	−0.69	−0.73
	(2.27)	(2.60)	(−3.39)	(−3.51)
Urban	−0.058	−0.058	−2.116	−2.103
	(−3.72)	(−3.83)	(−13.43)	(−13.30)
% Female (β_1)	0.142	0.030	−1.416	−1.482
	(2.36)	(0.36)	(−2.86)	(−2.96)
% Female*Law (β_4)	−0.105	0.055	−1.031	−0.723
	(−2.04)	(0.56)	(−0.98)	(−0.66)
Law (β_3)		−2.162		−0.205
		(−1.87)		(−0.97)
R-squared	0.613	0.642	0.027	0.028
Number of observations	49	49	7605	6605

Source: 1920 data from Goldin (1988, Table 1). 1880 rerun using 1880 Atack and Bateman (2004a). Numbers in parentheses are *t*-statistics

Notes: We have converted average daily scheduled hours of work in 1880 to a weekly basis by multiplying by 6 to put them on the same basis as those for 1920. In 1920s, Urban was equal to the % of population living in urban areas. In 1880, Urban is a dummy variable equal to 1 if a firm was located in a town or city with a population of 2500 or more

for women ($\beta_3 < 0$). In both, urbanization was associated with shorter hours while scheduled weekly hours of work in the South were substantially longer than those elsewhere.[58]

For 1880, we also conclude that the scheduled work-week was shorter in urban areas. Our estimates for 1880, however, are markedly different in a number of important respects.[59] First, our estimate of the coefficient on percentage of the firm's labor force that was female suggests that firms in those states without a maximum hours law where women were a larger proportion of the workforce had a shorter scheduled work-week $\beta_1 < 0$. Whereas Landes' concluded that women worked longer hours than men in those states where women were an important component of the labor force, we find that firms scheduled shorter hours. Second, although four states had passed a law regulating women's hours of work by 1880, hours of work in those states were not statistically different from those in states that had not passed such laws $\beta_3 = 0$. Third, women's hours were no different from men's hours in

[58]The means of the independent variables in 1920 were: SOUTH = 0.31; URBAN = 41.0; % Female = 12.3; LAW = 0.694.

[59]The means of the independent variables in 1880 were: SOUTH = 0.18; URBAN = 0.49; % Female = 0.049; LAW = 0.186.

those states that had passed these laws $\beta_4 = 0$. Taken together, these two results suggest that the laws may have had little or no effect. Fourth, the scheduled workweek was substantially shorter, not longer, in the South. This implies that sometime between 1880 and 1920 the southern states went from having shorter scheduled working hours to having longer scheduled working hours than other states. Indeed, one can make a strong case that overall the underlying structure in 1880 was almost completely different from that in 1920. In what follows, we focus solely upon 1880.

5.9 Impact of Hours Legislation in 1880

While child labor—ignored by Landes and Goldin—was of little and declining importance by 1920, it was by no means extinct as witness the 1918 U.S. Supreme Court decision in *HAMMER V. DAGENHART* declaring unconstitutional the Congressional prohibition of interstate commerce in goods produced by child labor.[60] In 1880, children were much more important. They made up more than 5% of the industrial labor force in our sample and in some industries—for example, furniture—were a seventh of the workforce (see Table 5.1). Consequently, with respect to 1880 at least the model in Table 5.5 is mis-specified. We must include the possibility that scheduled hours of work might be just as easily affected by limits on children's hours of work as by maximum hours laws for women. Moreover, despite their declaratory nature, it is possible that "legal day" laws may also have had an effect. We have also taken advantage of this opportunity to recognize that the effects may differ between male- and female-dominated industries. We use two different approaches. In one we created one dummy variable for male-dominated industries, defined as those industries where women made up less than 2% of the labor force and another dummy variable for female-dominated industries defined as those where women were at least 20% of the work force.[61] These results are shown in Table 5.6. In the other, we have used separate industry dummies with the textile industry as the omitted variable. This regression equation, with separate dummy variables for the Massachusetts maximum hours law for women, other states' maximum hours laws for women, legal day laws and child labor laws, is shown in Table 5.7.

The regression results for Specification (1) in Table 5.6 are similar to those for 1880 in Table 5.5. Shorter working days were associated with urban location and the southern states and women apparently worked a statistically significantly shorter day (about 12 min a day or an hour or so per week) than men in those industries whose workforce was 2–20% female. The coefficient on the dummy variable for state maximum hours law for women and the interactions between this

[60] 247 U.S. 251, 38 Supreme Court 529 (1918) declaring unconstitutional United States, C. 432, 64 Cong., 1 sess.

[61] Male-dominated industries: SIC codes 7, 17, 24, 35, 37, 38, 49 and 76; Female-dominated industries: SIC codes 21, 22, 23, 26, and 39. See Table 5.2 above.

Table 5.6 The impact of hours laws on the length of the scheduled workday, 1880

Variable	(1)	(2)	(3)	(4)
Constant	10.75	10.74	10.68	10.66
(H_m)	(414.70)	(413.57)	(333.76)	(331.62)
% Female	−0.195	−0.197	0.076	0.043
($H_f - H_m$)	(−2.09)	(−2.12)	(0.63)	(0.27)
% Children	−0.049	−0.054	−0.263	−0.261
($H_y - H_m$)	(−0.37)	(−0.41)	(−1.44)	(−1.42)
South	−0.145	−0.140	−0.084	−0.091
	(−4.19)	(−4.07)	(−2.20)	(2.39)
Urban	−0.343	−0.321	−0.328	−0.320
	(−12.85)	(−11.92)	(−12.17)	(−11.86)
Law_1 (Legal H_{fmax})	−0.050			
(rH_m)	(−1.41)			
Law_1*% Female	−0.091			
($qH_f - rH_m$)	(−0.49)			
Law_1*% Children	0.090			
($sH_y - rH_m$)	(0.29)			
Law_2(Legal H_{fmax} Mass.)		−0.311	−0.242	−0.408
(rH_m)		(−5.24)	(−3.85)	(−5.68)
Law_2*% Female		0.123	−0.141	−0.218
($qH_f - rH_m$)		(0.46)	(−0.48)	(−0.64)
Law_2*% Children		0.470	0.681	0.666
($sH_y - fH_M$)		(0.64)	(0.91)	(0.83)
Law_3(Legal H_{fmax} Not Mass.)		0.065	0.132	−0.031
(rH_m)		(1.59)	(2.90)	(−0.53)
Law_3*% Female		−0.043	−0.306	−0.384
($qH_f - rH_m$)		(−0.18)	(−1.13)	(−1.22)
Law_3*% Children		−0.084	0.131	0.115
($sH_y - rH_m$)		(−0.25)	(0.37)	(0.25)
Law_4 (Legal Day: H_{max})			0.110	0.055
(rH_m)			(3.33)	(1.57)
Law_4*% Female			−0.414	−0.406
($qH_f - rH_m$)			(−2.25)	(−2.17)
Law_4*% Children			0.441	0.422
($sH_y - rH_m$)			(1.69)	(1.47)
Law_5 (Child Labor Legal H_{ymax})				0.175
(rH_M)				(4.66)
Law_5*% Female				0.116
($qH_f - rH_m$)				(0.60)

(continued)

Table 5.6 (continued)

Variable	(1)	(2)	(3)	(4)
Law_5*% Children $(sH_y - rH_m)$				0.012 (0.04)
Male-dominated industries	−0.263 (−9.70)	−0.259 (−9.55)	−0.257 (−9.52)	−0.255 (−9.45)
Female-dominated industries	−0.301 (−6.54)	−0.307 (−6.68)	−0.314 (−6.83)	−0.320 (−6.98)
Adjusted R-squared	0.040	0.044	0.046	0.049
F-ratio	36.8	30.4	25.6	22.9

Source: Atack and Bateman (2004a)
Notes: Number of observations = 7630 in all specifications. *t*-statistics in parentheses

dummy variable and the proportion of each firm's workforce that was female or child are statistically insignificant and small. When we separated the Massachusetts law from those of the other three sampled states that had passed laws limiting the hours of work for women (Ohio, Minnesota and Wisconsin), the coefficient on the Massachusetts law was significant and negative implying that scheduled work hours in Massachusetts in these mixed industries averaged about 20 min less per day than in all other states. This finding proves robust across all specifications that separately identify Massachusetts. Introduction of a dummy variable for those states that had adopted a legal day (Specification 3) suggests that such laws were probably introduced in states that had longer hours and were ineffective in securing shorter hours for men. However, women in these states did have shorter scheduled work hours than men—about 25 min a day. It seems unlikely that this was a result of the law. More likely, the result reflects the industrial distribution of the female labor force and the customary hours in those industries. More importantly, the introduction of the legal day dummy variable erased any statistical significance for the coefficient on the percent of the firm's workforce that was female. This coefficient, it will be recalled, measures the divergence between male and female scheduled hours of work.

This result is robust with respect to the inclusion of child labor laws and the breakdown of the male- and female-dominated industry groupings into 2-digit SIC codes (Table 5.7). Two industries, SIC 20—food processing, a gender-mixed industry—and SIC 49—oil, gas and coke, a male-dominated industry—had significantly longer workdays than textiles and most other industries. Only one industry, printing (SIC 27) had significantly shorter working days.

We draw two important inferences from these regressions. First, the lack of statistical significance and the small coefficients attached to the proportion of the labor force that was female or child under all of the less restrictive equation specifications supports the contemporary assertions and evidence that men, women and children all worked the same hours. As a result, the determination of the labor force composition is independent of hours. Second, wherever Massachusetts is singled out, the coefficient on the legal dummy variable for Massachusetts is

consistently and statistically significantly negative and large. It may be argued that this simply reflects some unique, but non-specific, attribute of Massachusetts manufacturing. We think not.

Despite our reluctance to assert causal relationships, we believe that the quantitative and qualitative evidence support the assertion that the shorter working day in Massachusetts was the result of the passage of Massachusetts Ten-Hour law for women. Furthermore, consistent with the claim of the Ohio Commissioner of Labor, this legislation "being secured caused a lessening of the hours of adult male labor," while women reaped no differential specific gain. This is not entirely true, for as Fig. 5.1 suggests, the dominant employer of women in Massachusetts, the textile industry, had a longer working day than was customary in other industries prior to the passage of the Ten-Hour law, and therefore women gained the most.

Table 5.7 The impact of hours laws on the length of the scheduled workday, 1880

Variable	Coefficient	Variable	Coefficient
Constant	10.36	SIC20	0.645
(H_m)	(63.78)		(3.97)
% Female	0.027	SIC21	−0.191
$(H_f - H_m)$	(0.17)		(−1.10)
% Children	0.032	SIC23	−0.048
$(H_y - H_m)$	(0.17)		(−0.28)
South	−0.147	SIC24	0.102
	(−3.89)		(0.63)
Urban	−0.194	SIC25	−0.143
	(−6.85)		(−0.82)
Law_2(Legal H_{fmax} Mass.)	−0.371	SIC26	0.174
(rH_m)	(−5.27)		(0.87)
Law_2*% Female	−0.261	SIC27	−0.372
$(qH_f - rH_m)$	(0.78)		(−2.04)
Law_2*% Children	0.680	SIC28	0.205
$(sH_y - fH_M)$	(0.87)		(1.15)
Law_3(Legal H_{fmax} Not Mass.)	−0.013	SIC31	0.062
(rH_m)	(−0.22)		(0.38)
Law_3*% Female	−0.495	SIC32	−0.135
$(qH_f - rH_m)$	(−1.61)		(−0.77)
Law_3*% Children	−0.198	SIC33	−0.164
$(sH_y - rH_m)$	(−0.44)		(−0.96)
Law_4 (Legal Day: H_{max})	0.049	SIC34	−0.251
(rH_m)	(1.45)		(−1.27)
Law_4*% Female	−0.408	SIC35	−0.198

(continued)

Table 5.7 (continued)

Variable	Coefficient	Variable	Coefficient
$(qH_f - rH_m)$	(−2.23)		(−1.18)
Law_4*% Children	0.236	SIC36	−0.123
$(sH_y - rH_m)$	(0.84)		(−0.19)
Law_5 (Child Labor Legal H_{ymax})	0.170	SIC37	−0.304
(rH_M)	(4.63)		(−1.35)
Law_5*% Female	0.144	SC38	−0.206
$(qH_f - rH_m)$	(0.75)		(−0.50)
Law_5*% Children	−0.045	SIC39	−0.255
$(sH_y - rH_m)$	(−0.14)		(−1.46)
SIC7	−0.131	SIC49	0.549
	(−0.60)		(2.13)
SIC17	−0.184	SIC76	0.163
	(−1.12)		(0.99)
Adjusted R-squared			0.095
F-ratio			22.7

Source: 1880 Atack-Bateman sample
Notes: Number of observations = 7630 in all specifications. *T*-statistics in parentheses

5.10 The Impact of Hours Legislation on Labor Force Composition

Although a firm's hours do not seem to have been simultaneously determined with the labor force composition, the passage of laws limiting the hours of work for women and children may still have affected the labor force mix. Rather than comply with the law by reducing hours or violate the law, firms may have met its conditions by eliminating those workers whose employment would otherwise constrain the firm's scheduled work day. This is a plausible alternate scenario wherever and whenever states made efforts to enforce the laws. A rent-seeking model of the passage of hours laws for women and children would emphasize male support for these laws in the hope that employers would reduce employment of those workers affected by the law and increase demand for those who remained unconstrained. Again, we cannot rigorously test this proposition with cross-sectional data for a single year but we can see if the evidence is consistent with the hypothesis.

We have shown that in most industries occupations were segregated by gender. Men traditionally performed tasks that required physical strength and stamina, whereas women worked at those tasks that emphasized quickness and nimbleness. So too did children. Women and children may therefore have been regarded as closer substitutes for one another than men and women or men and children. Men could, however, be used in place of either. Indeed, the data on labor force composition in Table 5.1 above show that most firms in most industries employed no women or children. Nevertheless, there were some firms in many of the industries (except textiles, glass and brick, all metals and gas, coal and oil) that reported employing

only women or children or both, although these were clearly the exceptions. Excepting textiles, all of the industries in which no firms exclusively employed women or children were ones that placed a premium upon physical strength and endurance such as carrying bricks, stoking fires, and tapping furnaces.

Our model specifies that firms simultaneously determined the proportions of female and child workers they employed and that these proportions were influenced in part by laws governing the employment of each kind of labor. We have combined state compulsory education laws and child labor laws into a composite dummy variable representing constraints upon the employment of children. The impact of laws on labor force composition, however, depends critically upon whether or not any effort was made to enforce the laws. There is some evidence of at least haphazard enforcement of compulsory education and child labor laws. With respect to laws setting maximum hours for women, however, the only evidence of enforcement is from Massachusetts. We therefore restricted consideration to the Massachusetts 10-hour law.

Other factors also entered the decision such as traditional industry practices with respect to the employment of each kind of worker, the size of the labor force (as a proxy for the potential extent of the division of labor), the firm's capital-labor ratio (a proxy for the extent of mechanization), use of steam or water power (to capture the substitution of inanimate power for human muscle) and dummy variables representing the southern states and urban locations. The inclusion of the SOUTH dummy reflects Goldin and Sokoloff's (1982) argument regarding differential regional relative valuations for women and children compared with men in agricultural versus industrial pursuits. The URBAN dummy is to capture differential labor market conditions and opportunities between cities and rural areas.

More formally our model is[62]:

$$\alpha_f = \beta_{f0} + \varphi_{fy}\alpha_y + \beta_{f1}X_1 + \beta_{f2}X_2 + \ldots + \beta_{f10}X_{10} \tag{5.12}$$

$$\alpha_y = \beta_{y0} + \varphi_{yf}\alpha_f + \beta_{y2}X_2 + \beta_{y3}X_3 + \ldots + \beta_{y10}X_{10} + \beta_{y11}X_{11} \tag{5.13}$$

where X_1 = the Massachusetts 10-hour law; X_2 = number of employees; X_3 = Capital/Labor ($1000/person); X_4 = male-dominated industry; X_5 = female-dominated industry; X_6 = child-dominated industry; X_7 = use of waterpower; X_8 = use of steam power; X_9 = SOUTH; X_{10} = URBAN; and X_{11} = laws affecting the employment of children (includes both children's hours and compulsory education laws). Except for employment and the capital-labor ratio, all of the X_i variables are binary. In the model as specified the proportion of the workforce that was female does not directly depend upon the adoption of laws regulating the employment of children, nor does the proportion of children in the workforce depend directly upon laws limiting the hours of work for women.

The reduced forms of Eqs. (5.12) and (5.13) are:

[62]We are grateful to Alan Dye for deriving these equations and their reduced forms and for demonstrating that they are exactly identified.

$$\alpha_f = \pi f0 + \pi f1 X_1 + \ldots + \pi_{f11} X_{11} \tag{5.14}$$

$$\alpha_y = \pi_{y0} + \pi_{y1} X_1 + \ldots + \pi_{y11} X_{11} \tag{5.15}$$

where:

$$\begin{aligned} \pi_{f0} &= (\beta_{f0} + \varphi_{fy}\beta_{y0})/(1 - \varphi_{fy}\varphi_{yf}) \\ \pi_{f1} &= (\beta_{f1} + \varphi_{fy}\beta_{y1})/(1 - \varphi_{fy}\varphi_{yf}) \\ &\ldots \\ \pi_{f11} &= (\beta_{f11} + \varphi_{fy}\beta_{y11})/(1 - \varphi_{fy}\varphi_{yf}) \end{aligned} \tag{5.16}$$

Including the zero restrictions on β_{y1} and β_{f11}, this system is just identified and we can thus use the indirect least squares approach to estimate the parameters of the structural equations. The coefficients of the reduced form equations are first estimated using OLS and these unbiased and consistent estimates are then used to calculate the structural equation parameters. For example:

$$\pi_{y1}/\pi_{f1} = (\beta_{y1} + \varphi_{yf}\beta_{f1})/(\beta_{f1} + \varphi_{fy}\beta_{y1}) \tag{5.17}$$

Using the restriction that $\beta_{y1} = 0$,

$$\pi_{y1}/\pi_{f1} = \varphi_{yf}\beta_{f1}/\beta_{f1} = \varphi_{fy} \tag{5.18}$$

The resulting estimates are consistent but biased since they depend upon the ratio of the reduced form coefficients.

The coefficients of the reduced form equations are shown in Table 5.8. Most coefficients in each equation are statistically significantly different from zero at better than the 95% level and the equation to estimate the proportion of each firm's labor force that was female performed particularly well. Based upon these coefficients we have estimated the parameters of the structural equations. These are shown in Table 5.9.

As might be expected from the formulation of the structural equations, women and children appear as substitutes for one another. However, what was unexpected was the magnitude of the effect for women. The structural parameter φ_{fy} implies that a 1% point increase in the proportion of children employed is associated with a 2.6% point decline in the proportion of women in the workforce. Consequently, women stood to gain a substantial increase in their employment share from anything such as compulsory school attendance that caused the proportion of children a firm employed to decline. Were women therefore only concerned about the welfare of children when they agitated for compulsory school attendance and the passage of restrictions upon the employment of children? In contrast, a 1% point decrease in the proportion of women in the workforce, however, increased child employment by less than 1% points.

Table 5.8 The impact of maximum hours legislation in 1880 upon firm labor force mix

	Dependent variable: proportion of adult females in the firm's labor force	Dependent variable: proportion of males under 16 and females under 15 in the firm's labor force
Variable	Coefficient	Coefficient
Constant (π_0)	0.0304 (6.39)	0.0278 (8.17)
Massachusetts 10-hour law (π_1)	0.0274 (3.80)	−0.0179 (−3.49)
Number of employees (π_2)	0.0004 (12.65)	0.00005 (2.34)
Capital/labor (π_3)	−0.0043 (−4.07)	−0.0040 (−5.42)
Male-dominated industry (π_4)	−0.0249 (−6.80)	−0.0141 (−5.36)
Female-dominated industry (π_5)	0.2256 (35.61)	−0.0012 (−0.26)
Child-dominated industry (π_6)	−0.0815 (−11.93)	0.0384 (7.86)
Waterpower (π_7)	−0.0122 (−2.27)	−0.0060 (−1.56)
Steam power (π_8)	−0.0011 (−0.26)	0.0101 (3.29)
South (π_9)	−0.0071 (−1.38)	0.0050 (1.36)
Urban (π_{10})	0.0073 (1.97)	0.0143 (5.40)
Children's laws (π_{11})	0.0078 (1.84)	−0.0030 (0.98)
Adjusted R-squared	0.2305	0.0280
F-ratio	211.75	21.31

Source: Atack and Bateman (2004a)
Notes: Number of observations = 7630 in all specifications. *t*-statistics in parentheses

Laws restricting job opportunities for children through hours laws and compulsory schooling had the expected effect of reducing the proportion of children employed. However, its impact was very small. In states that had adopted such laws, the proportion of children employed was only 0.3% points lower than that in an otherwise identical firm operating in a state without such a law. The positive parameter for the Massachusetts 10-hour law for women, however, implies that the proportion of female employees was increased by the law. This seems unlikely although we note that organized labor had argued for shorter hours as a means of increasing employment. Instead, we believe that this probably reflects the greater

Table 5.9 Parameter estimates of the structural equations for the impact of enforced laws restricting employment on employment shares: 1880

	Dependent variable: proportion of adult females in the firm's labor force	Dependent variable: proportion of males under 16 and females under 15 in the firm's labor force
Variable	Coefficient	Coefficient
Constant (β_{i0})	−0.1464	−0.6830
Proportion of children (φ_{fy})	−2.6000	–
Proportion of women (φ_{yf})	–	−0.6533
Massachusetts 10-Hour law (β_{i1})	0.0274	–
Number of employees (β_{i2})	0.0016	0.0004
Capital/Labor (β_{i3})	0.0211	0.0098
Male-dominated industry (β_{i4})	0.0882	0.0435
Female-dominated industry (β_{i5})	0.3187	−0.2093
Child-dominated industry (β_{i6})	−0.0263	0.0212
Waterpower (β_{i7})	0.0398	0.0200
Steam power (β_{i8})	−0.0360	−0.0134
South (β_{i9})	−0.0085	−0.0005
Urban (β_{i10})	−0.0637	−0.0273
Children's laws (β_{i11})	–	−0.0030

Source: Computed from the reduced-form coefficients in Table 5.8

proportion of female workers in Massachusetts than elsewhere. That is, this dummy variable is picking up state-specific phenomena that may dominate the impact of the 10-hour law. Substituting a dummy for all states that had passed hours laws for women in place of the dummy representing the Massachusetts law yields a parameter estimate that is negative but very small (−0.00065) and not statistically significantly different from zero. It would appear, therefore, that women experienced no serious employment effects as the result of the passage of maximum hours laws for women although since the laws were thought to be ineffective everywhere except Massachusetts we may be simply measuring the degree to which those laws were ignored.

5.11 Compliance and Protective Legislation

Our discussion so far has focused on the extent to which businesses tried to comply with maximum hours laws and restrictions upon the employment of children by reducing hours or curtailing employment of the protected groups. There was, however, a third option—to break the law and risk prosecution and conviction. Some

firms—the Hamilton Manufacturing Company to name but one—preferred this to the alternatives.[63] Why? What factors entered that decision?

If the law was something less than a compelling moral force then compliance demanded enforcement. There had to be a mechanism for the detection and punishment of violators. In Massachusetts, the District Police were charged with enforcement and, as we have seen, did on occasion bring charges against those who ignored warnings. These resulted in conviction and fines. We have not yet located information on the prosecution of cases in the local courts of the other states that had passed laws limiting the hours of work for women or children but believe that such evidence of enforcement does exist. The contracting-out provisions of all but the Massachusetts law, however, made hours law enforcement for adults, difficult although not impossible. The Massachusetts Supreme Court, for example, had no difficulty deciding that the Hamilton Manufacturing Company had "willfully" violated the law. There is no reason why it should have been more difficult for other courts to decide whether coercion had been used. This was, after all, a concept well-established in criminal law. The enforcement of child labor laws, however, might have been more difficult where the courts demanded evidence that firms had "knowingly" violated the law. Such evidence presumably required the continued employment of a minor after the firm had been advised of the violation.

Where no penalties existed, compliance depended solely upon the moral force argument. Where fines could be levied, compliance was likely to be a function of the expected penalty. This depends upon the magnitude of the likely fine and the probability of detection. As we have argued, detection might be expected to be a function of firm location (urban areas were more likely to be policed and the existence of violations more widely known), industry (by custom or technological constraint, some industries had longer working days), the number of employees covered (the larger the number of protected employees the greater the likelihood of complaint and the greater the incentive for the authorities to seek compliance). In addition, we believe that two other, firm-specific factors might influence the decision: whether or not the firm used steam power and the size of the firm's fixed capital. Steam power involved fixed costs associated with firing the boilers and building up steam. The firm's capital investment is a proxy for the firm's fixed costs of production.

Whether or not a firm violated the law is, however, unobserved except in the few cases that were prosecuted. Instead we assume that scheduled hours in excess of the prescribed maximum and employment of the protected group are prima facie evidence of violation. This reliance upon the self-reporting of violations almost certainly biases the measure downwards. However, it seems preferable to one alternative we considered that classified as potential violators all those firms that

[63]For cases decided at the state supreme court level, see *COMMONWEALTH V. HAMILTON MANUFACTURING COMPANY*, 125 Mass. 383; *WENHAM V. STATE*, 65 Neb. 394, 400, 406; *STATE V. BUCHANAN*, 29 Wash. 602; *COMMONWEALTH V. BEATTY*, 15 Pa. Sup. Ct. 5, 17; *RICHIE v. PEOPLE*, 155 Ill. 98 (1895).

claimed to be just in compliance. We may be certain that some of those firms that employed covered groups and reported scheduled hours of work exactly equal to the maximum allowed violated the law. The objective evidence—scheduled hours of work—suggests that the law was probably a binding constraint upon them. They therefore had the incentive to violate it. As the Massachusetts District Police (1887, p. 55) observed:

> Occasional violations are detected, or reported to me, where there is temporary overwork to fill orders at the stipulated time, or to catch up in some department. In some cases, where the engine is started several minutes before the schedule (sic) time, complaints are made that the women, generally weavers, go to work five or ten minutes before the proper time, thus gaining one or two hours a week.

However, we cannot distinguish these firms from the others that not only claimed to comply but in fact did comply with law. Hence our preference for the self-reported violations measure despite the downward bias.

For women's maximum hours laws we have restricted consideration to the four sample states that had passed such law: Massachusetts, Minnesota, Ohio and Wisconsin and use samples of firms from the individual states (Atack and Bateman 2004b) rather than those firms from these states that appear in the 1880 national sample Atack and Bateman (2004a). The total sample size for these four states is consequently much larger: 2118 firms. Of these, 235 employed at least one women and hence could have violated the law. Of these, 139 reported scheduled hours that were exactly equal to the maximum permitted by their state laws for women. Perhaps more surprisingly, 54 firms (4 in Massachusetts, 2 in Minnesota, 13 in Ohio and 35 in Wisconsin), almost a quarter of the total, self-reported violations—that is, they reported scheduled hours of work in excess of the maximum permitted for women. In "male-dominated" industries such as lumber a majority of firms that employed at least one woman worked longer than the maximum permissible hours. The largest number of self-reported violations, however, was in the clothing trade where 14 of 56 firms reported longer hours than allowed by law. Most firms that employed women in Wisconsin which had the most restrictive law, 8 h per day, violated the law but presumably the women had "voluntarily" accepted contracts for longer hours. At the opposite extreme, only 4% of the Massachusetts firms employing women reported hours in excess of the legal maximum, consistent with the statutory prohibition on contracting-out and the apparently fairly rigorous enforcement of the law.

Logit estimates to determine what factors affected the probability that a firm would chose to violate the law are shown in Table 5.10. Two variables, the urban dummy and the Massachusetts dummy, were statistically significant at better than the 95% level and indicate that the probability of violation was lower in urban areas and lower in Massachusetts than elsewhere.

Two other dummy variables, use of steam power and "female-dominated" industry, were statistically significant at better than the 90% level. These also

Table 5.10 What influenced a firm's decision to violate maximum hours laws for women in 1880?

Variable	Coefficient
Maximum fine ($)	−0.00856
	(−0.434)
Urban dummy	−1.0023
	(−2.302)
Massachusetts dummy	−2.804
	(−4.703)
Number of female employees	−0.0133
	(−0.784)
Capital/labor ($1000 per employee)	0.5684
	(1.458)
Capital invested ($1000)	0.0186
	(0.608)
Capital invested squared	−0.0003
	(−0.779)
Steam power dummy	−1.1179
	(−1.839)
Female-dominated industry dummy	−0.7594
	(−1.687)
Constant	1.0043
	(0.862)
Number of observations	216
Chi-square (9)	71.78
Log likelihood	−78.52

Source: Calculated from Atack and Bateman (2004b). Logit estimates with *t*-statistics in parentheses. Dependent variable: 0 = in compliance; 1 = out of compliance

indicate that the probability of violation was lower in steam-powered plants and in female-dominated industries. Only the impact of steam power is contrary to that which we expected.[64]

As Table 5.11 indicates, these estimates do a good job of predicting whether or not a firm was out of compliance with their state law limiting the hours that women could be employed. We correctly identify 93% of those in compliance (157/168) as being in compliance and correctly predict 54% of the firms that were apparently out of compliance (26/48). Overall, our predictions were correct 85% of the time.

The estimates in Table 5.10 can be used to generate estimates of the probability of compliance or violation for firms with differing characteristics. These are calculated as:

[64]It seems unlikely that this is the result of multicolinearity with any of the obvious variables. For example, use of steam power was weakly negatively correlated with urban location (R = −0.064).

Table 5.11 Comparison of outcomes and probabilities from Table 5.10

Outcome	Pr < 0.5	Pr >= 0.5	Total
Failure	157	11	168
Success	22	26	48
Total	179	37	216

Table 5.12 Impact of firm characteristics and location upon the probability of violating maximum hours laws for women in 1880

Differential characteristic	Probability	Differential characteristic	Probability
No changes	0.167	Not "female dominated"	0.300
Not steam-powered	0.380	Massachusetts firm	0.012
Not urban	0.353	100 female employees	0.057
K/L = $2000/employee	0.261	Capital invested $100,000	0.057
Impact of different max fines			
Massachusetts-based firms			
Urban, $0 fine	0.018	Rural, $0 fine	0.048
Urban, $100 fine	0.008	Rural, $100 fine	0.022
Urban, $500	0.0003	Rural, $500 fine	0.0007
Non-Massachusetts-based firms			
Urban, $0 fine	0.235	Rural, $0 fine	0.048
Urban, $100 fine	0.116	Rural, $100 fine	0.263
Urban, $500	0.004	Rural, $500 fine	0.011

Source: Computed from Table 5.10 (see text for further discussion). Base characteristics: Maximum fine = $50; Urban = 1; Massachusetts = 0; 10 female employees; Capital/Labor = $1000/employee; Capital invested = $20,000; Steam power = 1; "Female-dominated industry" = 1

$$p(violated) = 1 - p(comply) = 1 - \frac{1}{[1 + e^{(b_0 + b_i X_i^*)}]} \tag{5.19}$$

for the set of characteristics defined by the vector X_i^*. Some estimates (an infinite number are possible) are shown in Table 5.12. The probability that an urban, steam-powered firm employing ten women in a female-dominated industry with an invested capital of $20,000 and a capital-labor ratio of $1000/employee in a state other than Massachusetts with a $50 fine for violation of the maximum hours law would violate the law was 0.167. In Massachusetts, the probability that this same firm would violate the law would have been only 0.012. On the other hand, if the same firm were located in a rural area outside Massachusetts, the probability that it would violate the law rises to 0.353. As an extreme case, a rural firm outside of Massachusetts that employed just one woman in a water-powered, non-female-dominated industry with a capital labor ratio of $500 and only $5000 invested capital—say a flour mill—would be expected to violate the law more than 70% of the time if the maximum fine was $50 and 80% of the time if there was no fine.

Although we cannot specifically identify the source of the greater compliance with the law in Massachusetts, we believe that it reflects the cut-and-dried nature of the law with its prohibition on contracting out. Even so, some firms would

still violate the law, especially in small rural areas and in industries not usually associated with female employment. In this respect, the Hamilton Manufacturing Company's decision to disregard the law was unusual though perhaps the successful state prosecution in that case helps explain our findings.

Unfortunately, the coefficient on the size of maximum fine that could be levied was not statistically significant. Perhaps this reflected the widespread irrelevance of the punishment so long as Massachusetts was the only state that enforced its law and levied the fine. Nevertheless, it had the expected sign—stiffer fines were associated with greater compliance. As a result, compliance in Massachusetts could have been virtually assured if the legislature had been willing to raise the maximum fine for violation to, say, $500. All this of course begs the question whether the legislature would have passed the law if the penalties had been that high and whether the courts would have been willing to levy the maximum fine even in cases of flagrant, willful and repeated violation. At the $500 level, however, only 3 urban firms in 10,000 and 7 rural firms in 10,000 would have broken the law. Even in those states that permitted contracting out, sharp increases in fines would probably have encouraged much greater compliance.

A similar logit model can be applied to the 1880 national sample to help illuminate the factors that aided or deterred compliance with the whole spectrum of hours laws including not only maximum hours laws for women but also child labor laws and the declaratory legal day laws (Table 5.13). On the whole though the results seem to provide much less insight into the impact of laws and the reaction of firms. Nationwide, about a quarter of all manufacturing firms in states that had passed child labor laws, adopted declaratory legal days or set maximum hours of work for women self-reported violation of the applicable law(s) if they required all workers to work the scheduled work day.

According to our estimates, larger, steam-driven firms in urban areas and in Massachusetts were much less likely to violate the laws than firms located elsewhere. By distinguishing individual industries at the two-digit SIC level (ferrous and non-ferrous metals are omitted) we can see that widespread violations of one or more of the laws were much less prevalent in only one industry, construction. It is interesting to note that this was one of two industries (the other being printing where violations were also less common) dominated by craft unions. On the other hand, violations were much more likely in food processing (SIC 20), clothing (SIC 23), lumber (SIC 24), paper (SIC 26), chemical (SIC 28), leather (SIC 31), brick and glass (SIC 32), gas, coke and oil (SIC 49) and blacksmithing (SIC 76). Except for the Massachusetts maximum hours law for women, the laws apparently had no statistical impact upon whether or not firms worked long hours. Notice, however, that the coefficient on the child labor laws was positive implying that firms employing children in states that only had a child labor law were somewhat more likely to be in violation of the law.

Overall, the model predicted correctly about 78% of the time (Table 5.14). As with the model examining the factors affecting compliance with maximum hours laws for women, the model predicted much better for compliance (94% correct) than violation (28%).

Table 5.13 Factors influencing a firm's decision to violate hours laws in 1880

Variable	Coefficient	*t*-Statistic
Maximum fine ($)	0.0043	1.361
Urban dummy	−1.0309	−10.259
Women's hours law (MA) dummy	−1.9662	−2.995
Women's hours law (not MA) dummy	−0.3396	−0.641
Legal day law dummy	−0.6875	−1.400
Child labor law dummy	0.4488	0.241
Persons affected by legal day law	−0.0013	−0.931
Persons affected by women's hours law	−0.0126	−0.895
Capital/labor ($1000 per employee)	−0.0091	−0.304
Capital invested ($1000)	0.0054	2.141
Capital invested squared	−5.11e−06	−1.323
Steam power dummy	−0.2960	−2.218
SIC 07 dummy	−0.3062	−0.284
SIC 17 dummy	−0.7720	−2.211
SIC 20 dummy	2.228	10.113
SIC 21 dummy	0.3321	0.957
SIC 22 dummy	0.4396	0.667
SIC 23 dummy	0.5613	1.938
SIC 24 dummy	0.779	3.334
SIC 25 dummy	0.3853	1.087
SIC 26 dummy	1.1012	2.984
SIC 27 dummy	−1.096	−1.461
SIC 28 dummy	1.208	3.819
SIC 31 dummy	0.7685	3.264
SIC 32 dummy	0.9174	2.622
SIC 35 dummy	−0.2743	−0.858
SIC 37 dummy	−0.4464	−0.582
SIC 38 dummy	1.4556	1.238
SIC 49 dummy	2.5251	3.872
SIC 76 dummy	0.9777	4.053
Constant	−0.9717	−1.820

Source: Calculated from Atack and Bateman (2004a). Dependent variable: 0 = in compliance; 1 = out of compliance. N = 2435. Chi-square (31) = 694.59 and log-likelihood = −1552.09

Table 5.14 Comparison of outcomes and probabilities from Table 5.13

Outcome	Pr < 0.5	Pr >= 0.5	Total
Failure	2441	164	2605
Success	597	233	830
Total	3038	397	3435

While the same exercise could be performed using the estimates in Table 5.13 to generate probabilities of violation as was done with Table 5.12, the results are not nearly so interesting. The probability of violation by a very large urban

Massachusetts textile mill with say $1 million invested and employing 500 women was virtually zero (4 in 10,000) but also very close to zero for a similarly large but non-urban textile mill in say Ohio. At the opposite extreme, a small rural flour mill in, say, upstate New York (which had adopted child labor laws and passed a declaratory legal day) would be expected to violate one or other of the laws about 28% of the time.

5.12 Conclusion

Protective legislation has now been under attack in Congress and the Courts for many decades. Our analysis of the early origins of this legislation, however, paints a quite different picture of motivation at its inception from current interpretations. Rather than being a symbol of the exercise of political power of men to obtain rents at the expense of women, the early laws, particularly those setting maximum hours for women, were proof of the inability of men to secure the shorter hours that many desired on their own either through industrial or legislative action. Instead, a coalition of social reformers and labor groups portrayed women as the fair and weaker sex in need of protection through shorter hours for the health and security of the nation. Had not women workers also sought shorter hours the interpretation would, of course, be quite different but there is ample documentary evidence that this ploy was adopted only after more direct methods had failed to achieve the desired result—shorter hours for everyone. Nor does the passage of early laws restricting women's hours of work appear to have adversely affected job opportunities for women.

For the twentieth century, Claudia Goldin's study suggests that protective legislation to limit hours had the effect of reducing hours of work for both men and women. In the nineteenth-century, this seems to have been true only in Massachusetts. Elsewhere the laws apparently had little or no effect. This conclusion suggests that enforcement played a crucial role. Only the Massachusetts law superseded contracts and prohibited contracting out prior to the late 1880s and only in Massachusetts does there seem to have been a rigorous and systematic effort to uncover and prosecute violations. The efficacy of this approach did not go unnoticed and the Massachusetts law became the model for similar laws in other states. Not all those laws, however, were found acceptable. For example, the Illinois Supreme Court struck down the Illinois eight-hour law for women in *RICHIE V. PEOPLE*[65] in 1895 while the US Supreme Court denied New York in *LOCHNER V. NEW YORK*[66] but upheld the state of Utah in *HOLDEN V. HARDY*.[67] The distinction between them seems to be that a clear case had to be established for the exercise of the

[65] *RICHIE V. PEOPLE*, 155 Ill. 98 (1895).

[66] *LOCHNER V. NEW YORK*, 198 US 45 (1905).

[67] *HOLDEN V. HARDY*, 169 US 366 (1898).

police powers of the state to protect health to supersede and "freedom of contract" under the 14th Amendment. This case was clearly made in the Massachusetts law and affirmed by the Massachusetts Supreme Court in *COMMONWEALTH V. HAMILTON MANUFACTURING COMPANY*[68] as it was in the other decisions affirming state laws at the federal level such as *MULLER V. OREGON*.[69]

No such case was made for the declaratory "legal day" laws. They seem to have been totally ineffective in reducing daily hours of work. Indeed, our results suggest that firms in states with such laws generally had longer scheduled hours of work than elsewhere. These laws stressed that workers could not be "compelled" to work longer hours but were free to contract to do so. However, one might argue, as many did, that the presentation of a contract for signature upon pain of dismissal constituted compulsion although the Courts never addressed the issue. Equally, employers were free to find a substitute work force that would be willing to work longer hours should the current work force be unwilling. Worker health was never a consideration.

Since children as minors had no freedom of contract, legal obstacles to restrictions on children's hours and work opportunities were minimal. Where laws were passed, they generally contained enforcement provisions and penalties and our results suggest that women may have benefited from increased employment opportunities where these laws were effective in removing some children from the factories. Many of the laws, however, were easily evaded especially where they declared that employers could not "knowingly" employ underage children for longer hours or children without the requisite educational attainment. Despite extensive policing efforts even Massachusetts seems to have found it harder to enforce the child labor laws in part because of the possibility of deception as to age and schooling by the parents or the children. Thus, companies might unwittingly violate the law. Carroll Wright in his survey of the impact of the Ten-Hour law in Massachusetts alluded to this in his diatribe against the French-Canadians in Massachusetts:

> They will not send their children to school if they can help it, but endeavor to crowd them into the mills at the earliest possible age. To do this they deceive about the age of their children with brazen effrontery. They also deceive about their schooling. (Massachusetts Bureau of Statistics of Labor 1880b, p. 470–71)

Within each firm, men women and children kept the same hours quite independent of any laws. This is consistent with the rigid sexual segregation of jobs common in the nineteenth century that resulted in men and women being complements rather than substitutes within the firm as well as with the desire by employers to monitor their workers, particularly in the larger plants. However, there were systematic variations in hours from industry to industry, between city and countryside and between North and South, reflecting forces such as custom, preferences and the demands of technology.

[68] *COMMONWEALTH V. HAMILTON MANUFACTURING COMPANY*, 120 Mass. 383 (1876).
[69] MULLER V. OREGON, 208 US 412, (1908).

Predictions about the characteristics of firms that might violate legal restrictions on employment generally accorded with expectations that larger, more visible firms such as those in urban areas or those employing large numbers of the affected group were more likely to be in compliance, particularly in Massachusetts. For laws regulating the employment of women, the evidence suggests, albeit very weakly, that the magnitude and certainty of penalties for violating the law may have been a major factor determining compliance. So much for law as a moral force!

Acknowledgements This paper was originally penned in 1987 and 1988 while the authors were Professor at the University of Illinois Urbana-Champaign and Visiting Professor at Harvard University and Professor at Indiana University respectively. Jeremy Atack moved to Vanderbilt University from Illinois in 1993. Fred Bateman moved to the University of Georgia in 1990. He died in 2012. In preparing the original draft, we benefited from perceptive comments by participants in seminars at the Harvard University, University of California at Los Angeles, University of Chicago, University of Illinois and Yale University Economic History Workshops and participants at 1988 Cliometric Conference. Among individuals making substantive comments and improvements were Lee Alston, Lou Cain, Alan Dye, Barry Eichengreen, David Galenson, Mary Eschelbach-Gregson (now Hansen), Larry Neal, Joseph Reid, Elyce Rotella, Andrew Seltzer, Kenneth Snowden, Richard Sutch, Peter Temin, Mark Toma, Thomas S. Ulen, David Weiman, Jeffrey Williamson and Robert Zevin. They are not responsible for any remaining errors or confusions and we apologize to anyone whose contribution we have missed. The principal change to the original version of this paper is to reflect the U.S. Supreme Court decision in UAW vs JOHNSON CONTROLS (499 U.S. 187 (1991)) that was handed down after the paper was originally written.

References

Atack J, Bateman F (1992) How long was the workday in 1880? J Econ Hist 52(1):129–160

Atack J, Bateman F (1999) Nineteenth-century US industrial development through the eyes of the census of manufactures: a new resource for historical research. Hist Methods J Quant Interdiscipl Hist 32(4):177–188

Atack J, Bateman F (2004a) National sample from the 1880 census of manufacturing: 2nd ICPSR version. Inter-university Consortium for Political and Social Research, Ann Arbor, MI

Atack J, Bateman F (2004b) State samples from the 1880 census of manufacturing: 2nd ICPSR version. Inter-university Consortium for Political and Social Research, Ann Arbor, MI

Baker EF (1925) Protective labor legislation. Columbia University, New York

Brandeis E (1918) Part 2: labor legislation. In: Commons J (ed) History of labor in the United States, 1896–1932, vol 3. Macmillan Company, New York

Cahill MC (1968) Shorter hours: a study of the movement since the civil war. Columbia University Press, New York

Commons JR (1918) History of labor in the United States, vol IV. The Machilolan Company, New York

Commons JR, Andrews JB (1927) Principles of labor legislation. Harper & Brothers, New York

De Graffenried C (1890) Child labor. Publ Am Econ Assoc 5(2):71–149

Derber M (1962) Testimony: the history of basic work hours and related benefit payments in the United States. In: Report and appendices, US presidential railroad commission, Washington DC, p Appendix 4

Edwards AM (1907) The labor legislation of Connecticut. Publ Am Econ Assoc 8(3):413–734

Goldin C (1988) Maximum hours legislation and female employment: a reassessment. J Polit Econ 96(1):189–205
Goldin C, Sokoloff K (1982) Women, children, and industrialization in the early republic. J Econ Hist 42(4):741–774
Goldmark JC (1912) Fatigue and efficiency. Russell Sage Foundation, New York
Illinois Bureau of Labor Statistics (1886) Fourth biennial report. HW Rokker, Springfield
Kelly MA (1950) Early federal regulation of hours of labor in the United States. Ind Labor Relat Rev 3(3):362–374
Kingsbury SM (ed) (1911) Labor laws and their enforcement: with special reference to Massachusetts. Longmans, Green, and Co., New York
Landes EM (1980) The effect of state maximum-hours laws on the employment of women in 1920. J Polit Econ 88(3):476–494
Landes WM, Solmon LC (1972) Compulsory schooling legislation. J Econ Hist 32(1):54–91
Massachusetts Bureau of Statistics of Labor (1872) Third annual report of the bureau of statistics of labor. Wright and Potter, Boston
Massachusetts Bureau of Statistics of Labor (1880a) Eleventh annual report of the bureau of statistics of labor. Rand, Avery & Co., Boston
Massachusetts Bureau of Statistics of Labor (1880b) Twelfth annual report of the bureau of statistics of labor. Rand, Avery & Co., Boston
Massachusetts District Police (1887) Report of the chief of the Massachusetts district police for the year ending December 31, 1886. Wright and Potter, Boston
Mitchell OG, Sorenson CT (1970) State labor legislation enacted in 1969. Mon Labor Rev 93:48–56
Montgomery D (1989) The fall of the house of labor. Cambridge University Press, New York
Ogburn WF (1912) Progress and uniformity in child-labor legislation. Columbia University, New York
Ohio Bureau of Labor Statistics (1879) Second annual report of the bureau of labor statistics . . . for the year 1878. Nevins & Myers, Columbus
Pennsylvania Secretary of Internal Affairs (1882) Annual report of the secretary of internal affairs of the commonwealth of Pennsylvania: part II, industrial statistics. Lane & Hunt, Harrisburg
Persons CE (1911) The early history of factory legislation in Massachusetts. In: Labor laws and their enforcement. Longmans, Green, New York, pp 123–125
Richardson JD (1908) A compilation of the messages and papers of the presidents, vol 3. Bureau of National Literature, Washington D.C.
Rosenbloom JL, Sundstrom W (1994) The decline in hours of work in us labour markets, 1890–1903. In: Grantham G, MacKinnon M (eds) Labour market evolution: the economic history of market integration, wage flexibility and the employment relation. Routledge, London, pp 161–84
Sokoloff KL (1986) Productivity growth in manufacturing during early industrialization: evidence from the American northeast, 1820–1860. In: Engerman SL, Gallman RE (eds) Long-term factors in American economic growth. University of Chicago Press, Chicago, pp 679–736
Sufrin SC, Sedgwick RC (1954) Labor law: development, administration, cases. Thomas Crowell, New York
Sumner H (1910) Report on condition of woman and child wage-earners in the United States: volume 9: history of women in industry in the United States. Government Printing Office, Washington DC
Towles JK (1908) Factory legislation of Rhode island. Am Econ Assoc Q 9(3):1–119
US Census Office (1872) 1870 census: volume 3. The statistics of wealth and industry of the United States. Government Printing Office, Washington DC
US Census Office (1928) Fourteenth census of the United States: manufactures, 1919, vol 8. Government Printing Office, Washington DC
US Congress, Senate (1910) Report on condition of woman and child wage-earners in the United States: volume 18: employment of women and children in selected industries. Government Printing Office, Washington DC

US Department of Commerce (1975) Historical statistics of the United States from colonial times to 1970. Government Printing Office, Washington DC
US Department of Labor (1896) Tenth annual report of the commissioner of labor. Government Printing Office, Washington DC
US Department of Labor (1904) Tenth special report of the commissioner of labor. Government Printing Office, Washington DC
US Department of Labor (1915) The boot and shoe industry in Massachusetts as a vocation for women. Government Printing Office, Washington DC
US Department of the Interior, Census Office (1883) Compendium of the tenth census. Government Printing Office, Washington DC
US Women's Bureau (1931) Chronological development of labor legislation for women in the United States. US Women's Bureau, Washington DC
Weeks JD (1886) Report on the statistics of wages in manufacturing industries: with supplementary reports on the average retail prices of necessaries of life, and on trades societies, and strikes and lockouts, vol 20. Government Printing Office, Washington DC
Whaples R (1990) Winning the eight-hour day, 1909–1919. J Econ Hist 50(2):393–406
Wright CD (1900) The history and growth of the United States census, vol 194. US Government Printing Office, Washington D.C.

Chapter 6
Political Selection of Federal Reserve Bank Cities

Jac C. Heckelman and John H. Wood

Abstract The Federal Reserve Act (1913) established the Reserve Board Organization Committee (RBOC) to determine the number and location of Federal Reserve districts and Reserve banks. Some scholars argue that the decisions were politically motivated but direct econometric evidence is lacking. A regression model utilizing solely political variables correctly predicts 11 of the 12 selected cities; the exception being Cleveland's selection over Cincinnati. Our results present direct evidence of the importance of political determinants for RBOC selection.

6.1 Introduction

> [I]t is the purpose of the committee, in each one of the cities visited, to ask the representatives of the bankers and businessmen who may attend, what their opinion is, not only with respect to their particular city and . . . the territory contiguous to that city which should be placed in the district, but also . . . what is their best judgment as to the number of districts into which the country as a whole should be divided.
> The question has to be approached in a very broad and patriotic spirit. . . . Of course, we all understand the very natural local pride which every community has and which prompts it to use every legitimate endeavor to secure one of the headquarters of these banks, but when you take into consideration that no matter where the headquarters bank is to be, there will be branch banks established in all important cities of the district, for that reason *the headquarters is not of such supreme importance as some people are inclined to believe.* When the system is fully organized and the branches are established in all the important localities, the facilities of the system will, of course, be readily accessible to all parts of the district. *I do not mean by that to say that the headquarters itself is not a matter of very great importance; it is a matter of the greatest importance.* The committee wants all the light it can get from bankers and businessmen and others interested, and it is for that reason that we have determined to have these hearings throughout the country.
> Secretary of the Treasury William G. McAdoo, chairman of the Reserve Bank Organization Committee (1914a), before taking testimony, New York City, Jan 5, 1914 (emphasis added)

J. C. Heckelman (✉) · J. H. Wood
Wake Forest University, Winston-Salem, NC, USA
e-mail: heckeljc@wfu.edu; jw@wfu.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_6

The origins of the Federal Reserve have attracted attention recently because of its centennial in 2013, as well as the remarkable innovations in its policies during and since the Great Recession of 2007–2009 (Bordo and Roberds 2013; Selgin et al. 2012; Romer and Romer 2013). Of special interest have been the determinants of the Fed's structure, how well that structure has served its purposes, and whether intellectual, technological, and geographic developments have made revisions desirable. Our paper addresses the first question.

The Federal Reserve Act of 1913 provided for a Reserve Bank Organization Committee (RBOC) to determine the number of districts (up to a maximum of twelve), district boundaries, and particular cities to host each district's reserve bank. The RBOC stated that its determination of reserve districts and cities gave "full consideration to the important factors," including (1) The ability of the member banks within the district to provide the capital required by law, (2) The mercantile, industrial, and financial connections existing in each district and the relations between the various portions of each district and its reserve city, (3) The probable ability of the Federal Reserve bank to meet the legitimate demands of business, whether normal or abnormal, in accordance with the spirit and provisions of the Federal Reserve Act, (4) The fair and equitable division of capital among districts, (5) The general geographical situation of the district, transportation lines, and communications facilities between the Reserve banks and all portions of the district, and (6) The growth of the district (Reserve Bank Organization Committee 1914b, pp. 3–4).

Perhaps reflecting the contradictions in Chairman McAdoo's statement above (particularly the italicized portions) regarding the importance of selected cities, some observers conclude the selections were economic (Odell and Weiman 1998; McAvoy 2006, 2012; Jaremski and Wheelock 2017), while others stress the importance of politics (Willis 1923; Miller and Genc 2002; Binder and Spindel 2013).

McAvoy (2006), Binder and Spindel (2013), and Jaremski and Wheelock (2017) conduct regression analyses on the cities which applied for Fed selection. McAvoy's political model performs poorly, incorrectly predicting several cities, and so he rejects the importance of any political effect. As explained in detail below, we believe his methodology is biased because it relied on only state-level data even though several states had multiple applicant cities. Binder and Spindel support a political influence model but their evidence is indirect. Their direct political measure is also state-level, and statistically insignificant. Their only evidence for political influence is that the coefficient on a regional dummy for the south is positive and significant, indicating that southern cities were more likely to be selected than cities from other regions. Yet, as they note, this could be either consistent with favoritism toward a region which was strongly Democratic, or a desire by the RBOC to counter the more limited credit in the South. Similarly, Jaremski and Wheelock's political variables are also state-level. We show local-level political variables are significantly correlated with city selection, lending greater support to a political influence model.

The selections made may have had unintended consequences. Specific city selection may not matter much for Fed policy *coordination*, but policy *preferences*

may differ among bank presidents and therefore policy outcomes may be affected by the location of the city in which the president resides. For example, bank presidents are expected to represent their regional (district) interests when voting on FOMC open market operations but they may be more attuned to local (city) economic conditions. Gildea (1992) explains that even though bank presidents are fully briefed on district conditions, they are bombarded by media stories which are "likely to be more local in nature", and finds that local variables better explain FOMC voting than do state or district level variables. Similarly, regional policy for Fed bank credit may depend on city or state conditions more so than overall district conditions when shocks are localized. This may have been especially important in the early days of the Fed when district information would have been less precise and slower to obtain. Local economic conditions near where the bank happened to be located may have been, in the face of uncertainty, extrapolated to represent the district as a whole. Thus, the decision on how Fed cities were selected may have had the unintended effect of affecting future monetary policy aside from any considerations regarding alleged efficiency of the new system. The problems of selection are therefore compounded if modern monetary policy is disproportionately directed by the needs of cities which were selected largely on the basis of political factors. We seek to determine to what extent RBOC city selection can indeed be predicted by political variables.

The paper is organized as follows. Section 6.2 identifies the functions of Reserve banks. Aside from personal preferences and politics, their locations should be those most productive of their purposes. Regression analysis in Sect. 6.3 indicates that political variables predict all but one of the cities selected by the RBOC. Specifically, our model predicts that Cincinnati better meets the political criteria consistent with selection of the other cities than does Cleveland, the city actually selected. The final section presents a summary.

6.2 Background

Decentralization was a condition of an American *central* (the very word was an epithet hurled at opponents' proposals) bank in 1913, although Congress left much of its application to a special committee and the Federal Reserve itself (Glass 1927, pp. 100, 167, 239; Wicker 2005, p. x). The Federal Reserve Act provided for a RBOC "to designate not less than eight nor more than twelve ... Federal reserve cities," to be located in corresponding districts "apportioned with due regard to the convenience and customary course of business and shall not necessarily be coterminous with any State or States. The districts thus created may be readjusted and new districts may from time to time be created by the Federal Reserve Board, not to exceed twelve in all. The determination [of the RBOC] shall not be subject to review except by the Federal Reserve Board when organized" (Sec. 2).

The Board consisted of the Secretary of the Treasury and the Comptroller of the Currency, ex officio, and five members appointed by the president for rotating

ten-year terms (Sec. 10). Federal Reserve credit was the joint responsibility of the Board and the Reserve Banks. The Fed's main tools of monetary policy were initially the Reserve Bank discount rates, chosen subject to the Board's approval, at which it lent to member banks. Since World War II, Fed credit has overwhelmingly consisted of purchases of government securities as decided by the Federal Open Market Committee (FOMC) composed of the Board and a minority of Reserve Bank presidents.[1] The RBOC consisted of the Secretary of the Treasury in the chair (William G. McAdoo), the Secretary of Agriculture (David F. Houston), and the Comptroller of the Currency (J. Skelton Williams), all southern democrats, one of whom, the chairman, was politically ambitious.

The number and locations of Federal Reserve districts and banks were controversial. One proposal included forty-eight (one per state), or more if necessary to maintain personal relations between Reserve and member bankers. Any bank was to be no more distant than an overnight's train ride so that the representative of a bank threatened by a run could take commercial paper by train to the Reserve city and wire that he had cashed sufficient securities to meet depositors' demands. The Republicans' United Reserve Bank and Aldrich plans had 20 and 15 branches, respectively. In contrast, New York banker Paul Warburg testified to the RBOC that eight Reserve banks were too many, believing that for administrative purposes they should be few and close to New York and Washington. The November 1915 report of the Federal Reserve Board Committee on Redistricting proposed reducing the number of Reserve Banks from 12 to 8 or 9, but the attorney-general gave the opinion that this was beyond the powers of the Fed. Others argued that a few large Reserve banks would limit the power of New York, which would dominate a system of many small districts. Carter Glass's (Chair of the House Banking and Currency Committee) initial proposal mandated 20 reserve banks. This was changed in committee to a minimum of 12 and a maximum of 20. The Senate wanted fewer and the conference committee settled on a maximum of 12 and a minimum of 8. As indicated above, the final Act assumed the RBOC would choose fewer than the maximum (Timberlake 1978, p. 193; Weyforth 1933, pp. 18–20; Warburg 1930, pp. 759–791).

Secretary Houston (1926, p. 103) recalled the RBOC's "difficulty in arranging more than eight or nine strong districts and yet I felt that, since the law permitted twelve, it might be better . . . to fix that number and be done with it." This accorded with the view of Secretary McAdoo (1931, p. 258), who "saw early in our investigation that twelve Reserve Banks [were] better than eight." "There can be little doubt," McAdoo's biographer wrote, "that political exigencies dictated the creation of a full complement of twelve districts . . .; to establish fewer would

[1]The Federal Reserve Act of 1913 provided that the governor and vice governor were to be appointed by the president from among the Board's members. The Banking Act of 1935 revised the Board by dropping the Secretary and the Comptroller, and providing for seven members, all called governors and appointed for 14-year terms. It also formalized the FOMC, with a majority from the Board (seven, with five Reserve Bank presidents), but otherwise left the structure of the Federal Reserve unchanged.

have split the party.... The final decision on number and locations was apparently made at a conference" of RBOC members and Colonel Edward House (of Texas), Wilson's chief political advisor (Broesamle 1973, p. 120).

Except for the first of the six criteria the RBOC gave for its determining factors,[2] none of the RBOC's criteria was specific, and it was accused of rationalizing selections by inconsistent criteria.[3] Several selections were criticized for apparent personal or political influence. Missouri, the only state with two Reserve cities, had two Democratic senators; five of the twelve Reserve cities were located in the solid (Democratic) south, with less than a third of the nation's population (and falling); Richmond, with obvious personal connections, was not thought a strong candidate before it launched a vigorous campaign and pulled influence away from the northeast (Willis 1923; Binder and Spindel 2013).[4]

A competing explanation was based on economics, especially geography and communications, as indicated by factors 2, 5, and 6 (McAvoy 2006; Jaremski and Wheelock 2017), although their contributions to monetary policy were not made clear. The RBOC asked bankers about the extents to which their preferred cities were "pre-eminently the business and financial center of the district as you have laid it out," although many banks had closer relations with distant money-center correspondents than with candidate Reserve cities in what became their Fed districts.[5] The New York Reserve bank's Benjamin Strong later stated that monetary policy decisions were based on the portfolios of New York and Chicago commercial banks.[6]

Bank appeals from the RBOC's decisions indicated their preferences for existing relations: banks in northern New Jersey and western Connecticut asked to move from the Philadelphia and Boston districts to the New York district; northern-Wisconsin and Michigan upper-peninsula banks from the Minneapolis to the Chicago district; Oklahoma banks from the Dallas to the Kansas City district; northwestern West Virginia banks from the Richmond to the Cleveland (Pittsburgh branch) district; and banks in southern Louisiana from the Dallas to the Atlanta

[2]The first action, from Section 2 of the Federal Reserve Act, required every national bank to subscribe to the capital stock of its Federal Reserve bank a sum equal to 6% of its own capital.

[3]C.H. Sloan, Republican congressman from Nebraska said the process lacked systematic, objective selection criteria. In some instances "the size of the cities did not control"; similarly with "the lines of transportation and the course of business, population, and banking preferences." "I do not know a single basis or rule that controlled in one case which has controlled in another. [P]olitics [favoritism] must have had something to do with it" (U.S. Congress 1914, p. 6441). See also McAvoy (2006).

[4]It has been noted that Willis' 1923 explanation of location choices emphasized politics more than his 1914 explanation, written while he was employed by the Federal Reserve (Hammes 2001).

[5]On New York and Cincinnati, see Reserve Bank Organization Committee (1914a, pp. 83, 4360–4364).

[6]The "Strong rule" is discussed by Chandler (1958, p. 240).

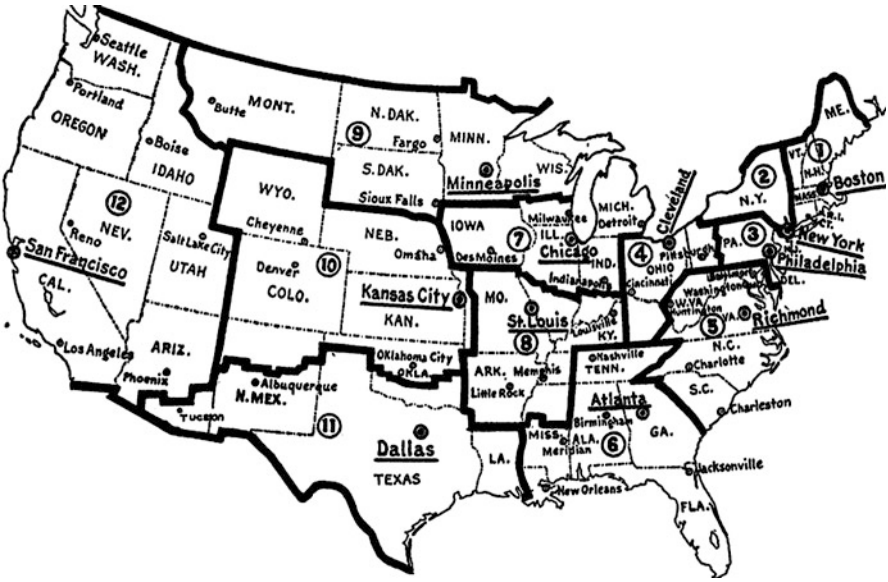

Fig. 6.1 Map of original (1913) Federal Reserve districts and cities. Source: Reserve Bank Organization Committee (1914b)

(New Orleans branch) district (see Fig. 6.1). Most requests were granted. Denied were those from Michigan's upper-peninsula and Nebraska and Wyoming to move to the Chicago district (Federal Reserve Board 1915).

Moving required reserves to the new Reserve banks did not mean new working relations. The coming of the Federal Reserve did not reduce correspondent banking between money-center and interior banks (Brown 1930; Watkins 1929). A lender of last resort might have worked through correspondent relations independent of the necessarily artificial Fed districts.

The concept of Optimal Currency Areas has also been discussed (Beckworth 2010; Rockoff 2003; Miller and Genc 2002). Congressman Carter Glass stated in his presentation of the Federal Reserve bill that "Each member bank is to deal directly with its regional reserve bank in securing rediscounts .... While subject to limited control by the Federal Reserve Board, the regional reserve bank is given an independent status as well as exceedingly important functions. It has the initiative in fixing rates of discount within its territory and the exclusive determination of the amount of paper to be rediscounted for member banks" (U.S. Congress 1913, p. 4645). Yet, the regional independence of monetary policies is limited in a common currency (dollar) area. The impact of easy money in a Fed district is diffused over a wider area and nineteenth-century panics had nationwide effects (Sprague 1910). Correspondent banking that made for a national currency and credit

system was well-developed in the nineteenth century, as were clearinghouses, and the Fed's bank regulation was a redundant addition to existing state and national regulations (Watkins 1929; Brown 1930; White 2013).

Legislators and the RBOC were not constrained by economic relations, however, and our interest is in their reasons for Reserve bank locations. Selection of Federal Reserve cities was clearly more political than economic. Actually, it would seem that none of the alleged economic factors is a persuasive reason for the Fed bank locations. The size and recent growth of a city, its mercantile, industrial, and financial development, and its transportation and communications lines might be correlated with financial flows, but they are not determining. The cure for an inelastic currency, as Thornton (1802) and Bagehot (1873) wrote, is the assurance that cash is and will continue to be accessible. This is more than bailing out specific banks, which is definitely not a solution. Solution of a cash shortage (usually caused by hoarding because of fears of a future shortage) means getting money into the pipeline, along with the assurance that there is more from where that came. Such a policy was already available through correspondent networks that connected the money centers with each other and the rest of the country. Putting a Federal Reserve Bank in a city with correspondent relations (meaning everywhere) was redundant, although it might have been thought that increases in cash might be more assured by the Federal Reserve than the profit-constrained Bank of England. In fact, the reverse occurred as the Fed became more interested in the equity positions of particular banks than of cash in the system (Schwartz 1992; McKinley 2012; Wood 2015). Bank desires for nearby Fed banks might have had their origins in the hope of such special treatment.

Membership in the Federal Reserve System was mandatory for the nearly 7500 national banks, but they could get around this requirement by switching their charters to join the 20,000 state-chartered banks for whom participation was voluntary. In practice, most banks preferred state charters, and most of those eschewed Fed membership (choices which still prevail). The selection of particular cities, therefore, virtually required a political (popular) solution dependent on bank preferences to entice membership in the new system.

Harold Reed (1922, p. 18) called the federal system a "politically clever" way to secure a central bank. Since its very conception and every succeeding step in the Fed's development had involved political compromise, particularly between populists and money center banks, it would have been surprising if Reserve bank locations had not been more of the same (Kolko 2008, pp. 217–254; Wood 2005, pp. 156–166). Reserve cities could be chosen simply on political grounds because, given communications technology and interbank relations, even in 1913, the effectiveness of monetary policy was independent of the economic characteristics of cities, particularly in light of branches which performed many of the same functions (especially check-clearing) and had many of the same powers (such as rediscounting) as the headquarters Reserve banks. Specific policy preferences, however, are highly dependent on localized city conditions (Gildea 1992).

6.3 Regression Analysis

McAvoy (2006) represents the first quantitative attempt (specifically, regression analysis) to study the determinants of the RBOC's selections. Related work was followed by Binder and Spindel (2013) and Jaremski and Wheelock (2017). McAvoy creates a political model and when it performs poorly, rejects it in favor of an economic model. We take issue with his interpretation and show his methodology is biased against the political model. Our alterations to the political model suggest it performs much better than McAvoy allows, and successfully predicts 11 of the 12 selected cities, except for the choice of Cleveland over Cincinnati. This is somewhat ironic in that the choice of Cleveland is widely understood to be political in nature (Miller and Genc 2002).

McAvoy (2006) considers the 37 cities that applied for Reserve Bank status. A dummy dependent variable is created to indicate which of the sample cities were selected. McAvoy then develops a "political model" by arguing that the political motivation would be to award locations based on partisan influence from the Democratic Party or from the House or Senate Banking Committee. These are measured by the number of senators from the city's state that are Democrats, the number of representatives from the city's state that are Democrats, and an indicator variable for whether or not the city's state had representation on either the House or Senate Banking Committees. Data values for these variables are presented in Table 6.1. The 12 selected cities are in bold, and assigned the value of one for the dependent variable; the remaining 25 cities are assigned the value of zero. Note that some states, in particular Ohio, Pennsylvania, Georgia, Alabama, Missouri, Texas, and Washington, had multiple cities petition. Only Missouri became a state with multiple cities selected.

Table 6.2 presents marginal impact estimates from logit analysis using robust standard errors. In column (I) we replicate McAvoy's model. Estimates differ slightly but the overall interpretation is similar. As found by McAvoy, the overall fit is poor (measured here by the McFadden R-Squared) and only the House variable is even marginally significant. Predicted values of selecting each city are presented in Table 6.3. There are several "incorrect" cities in the top 12; Cincinnati, Columbus, Pittsburgh, Fort Worth, and Houston are included while the true selections of San Francisco, Minneapolis, Atlanta, and Richmond are left out. The list of predicted selected cities matches McAvoy except his estimates also failed to include Boston. Based on these results, McAvoy rejects the political model.

There is a serious flaw in this methodology. The specification only includes state level variables to predict selected cities. Because several states have multiple cities in the sample, the model is misspecified. Note that every city incorrectly predicted to be selected is from the same state in which a city was correctly predicted to be selected. In other words, in order to correctly predict the selection of Cleveland, the other Ohio cities of Cincinnati and Columbus must also be predicted to be selected. The same is true for Fort Worth and Houston due to the correct prediction of Dallas being selected. Furthermore, in order to predict correctly that St. Paul would not

Table 6.1 Data

City	State	District	Senate	House	Banking	Votes
Boston	**MA**	1	0	8	1	291
New York	**NY**	2	1	31	1	673
Philadelphia	**PA**	3	0	12	1	509
Pittsburgh	**PA**	4	0	12	1	355
Cincinnati	**OH**	4	1	19	1	301
Cleveland	**OH**	4	1	19	1	112
Columbus	**OH**	4	1	19	1	64
Wheeling	WV	4	1	2	0	0
Washington	DC	5	0	0	0	28
Baltimore	MD	5	2	6	0	141
Richmond	**VA**	5	2	9	0	170
Charlotte	NC	5	2	10	0	19
Columbia	SC	5	2	7	1	0
New Orleans	LA	6	2	7	0	51
Atlanta	**GA**	6	2	12	0	124
Savannah	**GA**	6	2	12	0	24
Birmingham	AL	6	2	10	0	55
Montgomery	AL	6	2	10	0	4
Chattanooga	TN	6	2	8	0	11
Chicago	**IL**	7	1	20	1	908
St. Louis	**MO**	8	2	14	1	302
Louisville	KY	8	1	9	0	116
Memphis	TN	8	2	8	0	16
Minneapolis	**MN**	9	0	1	1	413
St. Paul	**MN**	9	0	1	1	95
Denver	CO	10	2	4	1	136
Kansas City	**MO**	10	2	14	1	497
Omaha	NE	10	1	3	1	220
Lincoln	NE	10	1	3	1	22
Dallas	**TX**	11	2	18	1	247
Houston	**TX**	11	2	18	1	97
Fort Worth	**TX**	11	2	18	1	85
San Francisco	**CA**	12	0	3	1	259
Seattle	WA	12	0	0	0	40
Portland	OR	12	2	0	0	75
Salt Lake City	UT	12	0	0	0	31
Spokane	WA	12	0	0	0	30

Note: Selected cities, and the states they are in, are bolded. Senate is the number of Democrats representing the state in the Senate and House is the number of Democrats from the state in the House of Representatives; Banking is a dummy variable indicating if the state is represented on the Senate Banking Committee; Votes indicates the number of first place votes from a poll of national banks

Table 6.2 Logit analysis for selected cities

	(I)	(II)	(III)	(IV)
Constant	−2.03	−3.35	−42.59	−116.06
	(−2.49)	−(2.37)	(−2.02)	(−2.55)
Democratic senators	−0.096	−0.515	−0.008	1.80e−5
	(−0.98)	(−1.83)	(−0.58)	(2.06)
Democratic representatives	0.022	0.139	0.0008	7.65e−6
	(1.82)	(2.39)	(0.44)	(2.87)
Banking committee	0.240	0.470	−0.010	0.002
	(1.30)	(2.01)	(−0.68)	(3.54)
First place votes			0.052	2.38e−5
			(1.87)	(3.58)
County votes for Wilson			0.062	0.0001
			(1.94)	(2.10)
Predict	City	State	City	City
District dummies?	No	No	No	Yes
N	37	25	37	33
Mean, Dependent variable	0.324	0.440	0.324	0.265
McFadden R-Squared	0.202	0.646	0.637	0.778

Note: Slopes estimate at means (except for constant) with robust z-statistics in parentheses

be selected the model forces Minneapolis to be incorrectly predicted not to be selected. Similarly, Atlanta cannot be correctly predicted to be selected if Savannah is correctly predicted not to be selected. Because the model cannot distinguish cities in the same state, it cannot accurately predict the 12 selections. McAvoy's political model fails by construction.[7]

Because the political model relies exclusively on state-level data, it is better suited to predict which states receive a Federal Reserve bank, rather than which particular cities are selected. McAvoy's cursory description of possible political motivations, found in his footnote 6, explains the potential selection of various cities based on a prominent Democrat being from that state. Yet if this were the true rationale, then it would not matter which particular city in the state received the preferential treatment. For example, McAvoy explains that President Wilson being from Virginia is one reason critics point to the selection of Richmond. We acknowledge that because Richmond was the only applicant city from Virginia, Richmond must be selected in order to reward Virginia. But this does not explain other city selections. In the same footnote, McAvoy also notes the critique that "Secretary Houston was formerly president of Texas A&M University, so Dallas received a FRB." Yet this reasoning can only explain why Texas received a Federal Reserve Bank city, not specifically why it was Dallas. In fact, Texas A&M

[7]In contrast, the "economic model" McAvoy promotes uses mainly city and county-specific independent variables, as well as twice as many total variables.

Table 6.3 Predicted cities from McAvoy's model

City	Probability	Rank
New York*	0.887	1
Chicago*	0.706	2
Cincinnati*	0.683	3
Cleveland*	0.683	3
Columbus*	0.683	3
Philadelphia*	0.620	6
Pittsburgh	0.620	6
Dallas*	0.547	8
Fort Worth*	0.547	8
Houston*	0.547	8
Boston*	0.515	11
Kansas City*	0.439	12
St. Louis*	0.439	12
San Francisco*	0.382	14
Minneapolis*	0.333	15
St. Paul*	0.333	15
Omaha	0.278	17
Lincoln	0.278	17
Columbia*	0.269	19
Denver	0.211	20
Louisville	0.117	21
Atlanta*	0.156	22
Savannah*	0.156	22
Charlotte	0.130	24
Birmingham	0.130	24
Montgomery	0.130	24
Richmond	0.118	27
Seattle	0.116	28
Spokane	0.116	28
Salt Lake City	0.116	28
Washington	0.116	28
Chattanooga	0.107	32
Memphis	0.107	32
New Orleans	0.097	34
Wheeling	0.092	35
Baltimore	0.088	36
Portland	0.048	37

Note: * indicates highest probability in the district; italics indicates actual reserve bank city with less than the highest probability in its district. Probabilities based on column (I) of Table 6.2

University is located closer to the applicant city of Houston than to Dallas (or Fort Worth). Similarly, McAvoy suggests that one reason critics give for Atlanta being selected is because Chairman McAdoo was born in Georgia. Yet, Savannah could also have been selected for this purpose. Thus, McAvoy's presented political motivations (which he wishes to dismiss) are better suited for predicting states than cities.

The 37 applicant cities represent 25 distinct states. To test the state political model, we create a new binary dependent variable indicating whether that state contained a selected Federal Reserve city. (These are bolded in the second column of Table 6.1.) Logit estimates using the state dependent variable are presented in column (II) of Table 6.2.

The state model performs much better than the city model. The number of Democrats in the House representing a district in that state and someone from the state on the Senate or House Banking committees are statistically significant at better than 1%, and the overall model fit increases by more than three-fold. As shown in Table 6.4, 10 of the 11 states are correctly predicted,[8] with only Kentucky being incorrectly predicted instead of Virginia. The probability of Kentucky being selected, though, is below that of Missouri which was also predicted, and Kentucky's lone city applicant of Louisville is in the same district (8) as St Louis. No state is predicted to be among the top 11 for district 5. The highest probability, though, for any state within the fifth district is South Carolina, rather than the actual selected bank in Virginia (Richmond). Thus district 5 is still incorrectly predicted even on a district-by-district selection basis. Note that in the state model, Ohio is correctly predicted for district 4 as the distinction between Cleveland versus Cincinnati or Columbus becomes irrelevant. Indeed, if the political model is designed to test if the RBOC was influenced by Democrats from Ohio (or any other states) then rewarding Ohio occurs regardless of which Ohio city is selected.

Ultimately, though, we wish to predict selected cities rather than just the states. We therefore need to supplement McAvoy's political model specification with local level variables, to be able to distinguish among cities within a given state. We do this two ways. First, we consider the RBOC's desire to reward a city based on support for the Democratic Party by including the logged county vote share for Woodrow Wilson in the 1912 presidential election.[9] Second, we consider political influence from the bankers by including the logged number of votes a city received. To gauge bank preferences, two polls had been taken of the 7471 national banks which would become part of the new Federal Reserve System (Gerena 2007).[10] First, banks were instructed to list from 8 to 12 potential cities for selection. The

[8]Recall that both Kansas City and St. Louis are Federal Reserve cities, so the 12 cities represent 11 different states.

[9]County votes are taken from Dave Leip's *Atlas of U.S. Presidential Elections*. In some cases, cities are not part of any county and the votes represent the city total exclusively. In all other cases, county level is the lowest level of aggregation available.

[10]The Federal Reserve Act made Fed membership of the more than 19,000 state banks voluntary, and few initially joined (Federal Reserve Board 1959, pp. 41–45).

Table 6.4 Predicted states

State	Probability	Rank
NY*	1.000	1
IL*	0.999	2
OH*	0.999	3
PA*	0.999	4
TX	0.999	5
MA*	0.997	6
MO*	0.991	7
CA*	0.861	8
KY	0.695	9
GA*	0.568	10
MN*	0.562	11
SC*	0.299	12
NE	0.252	13
AL	0.214	14
NC	0.214	14
VA	0.110	16
TN	0.053	17
CO	0.039	18
DC	0.034	19
UT	0.034	19
WA	0.034	19
LA	0.025	22
MD	0.011	23
WV	0.009	24
OR	<0.0001	25

Note: * indicates highest probability in the district; italics indicates state with actual reserve bank city with less than the highest probability in its district. Probabilities based on column (II) of Table 6.2

range listed matches the language in the Federal Reserve Bank Act because the RBOC had not yet determined how many districts would be created. Second, banks were told to rank their first, second, and third preferences for the city with which it would be affiliated. These polls occurred prior to any cities officially applying for Federal Reserve Bank city status.

The RBOC did not specify how they would tally the preference votes but the Reserve Bank Organization Committee (1914c) presents a separate tabulation of just first place votes from the second poll stratified by district location of the voting bank, ignoring entirely the second and third place votes. Furthermore, Elliott's (1914)

report only lists vote totals for the top 29 vote-receiving cities from the first poll.[11] We therefore consider city popularity by the number of first place votes from the second poll. Vote totals for our sample cities are presented in the last column of Table 6.1.

Although McAvoy attributes bank preferences as representing social efficiency, there are many reasons a bank may prefer one city over another unrelated to efficiency. We perceive banks to act in their own self-interest, and vote accordingly. Thus banker preferences represent a special interest, and we treat bank votes as another political variable, although distinct from partisan or legislator preferences. Private bankers have their own regulatory preferences (interests) which can differ from the social interest, and various studies have presented evidence of their influence on laws, reserve requirements, and other regulations (Havrilesky 1995; Lown and Wood 2003; Mason 2004; Heckelman and Wood 2008). They may have also influenced the selection of cities that would be home to their regional regulators. To the extent that those to be regulated influenced the choice of regulator, we consider this political influence, although we note that Jaremski and Wheelock (2017) classify bank preferences as separate from either economic or political variables.

Logit estimates from the local-level political model using the 37 city sample are presented in column (III) of Table 6.2, and corresponding city probabilities are presented in Table 6.5. Both new variables have the predicted positive sign and are borderline significant. Their inclusion more than triples the regression fit (relative to McAvoy's original city prediction model in column (I)) and reduces, or altogether eliminates, significance of the state variables, further indicating a misspecification from the original model. Examining the predicted selected cities in the first set of columns of Table 6.5 reveals that while Dallas remains correctly predicted to be selected, this no longer forces a predicted selection for either Houston or Fort Worth, both of which now have probabilities below 30%. However, although Columbus is no longer incorrectly predicted (its new independent prediction level is 0.006), Cincinnati remains incorrectly predicted to be selected, and with greater probability of selection than Cleveland. Overall, the model correctly predicts 10 of the 12 Federal Reserve cities, a significant improvement over McAvoy's original model which used only state level data. Richmond jumps to over a 90% probability of selection, and Minneapolis also climbs to a roughly 75% chance of selection and is now safely in the top 12. However, St. Louis and Cleveland fall out, with Cleveland being an especially surprising selection under this model, with only a 7.6% probability of selection.

In defending its decisions to Congress, the RBOC compared selected cities against alternative cities in the same district. This would tend to suggest that

[11] Los Angeles received the 25th highest total but ultimately did not apply to become a Reserve Bank city and is not part of our sample. Seattle was the last listed city with 98 votes. Thus the remaining nine applicant cities in our sample could have received anywhere from zero to 97 votes. None of them were selected by the RBOC.

Table 6.5 Predicted cities when including 1912 presidential election votes and bank votes

No fixed effect included			District fixed effect included		
City	Probability	Rank	City	Probability	Rank
New York*	0.999	1	Boston*	Automatic	1
Chicago*	0.993	2	New York*	Automatic	1
Kansas City*	0.982	3	Philadelphia*	Automatic	1
Dallas*	0.966	4	Chicago*	Automatic	1
Richmond*	0.920	5	Kansas City*	0.999	5
Boston*	0.892	6	San Francisco*	0.999	6
Philadelphia*	0.855	7	St. Louis*	0.999	7
San Francisco*	0.803	8	Richmond*	0.988	8
Cincinnati*	0.798	9	Atlanta*	0.890	9
Minneapolis*	0.758	10	Dallas*	0.843	10
Atlanta*	0.678	11	Cincinnati*	0.756	11
Pittsburgh	0.475	12	Minneapolis*	0.478	12
St. Louis*	0.454	13	St. Paul	0.252	13
Omaha	0.392	14	*Cleveland*	0.233	14
Fort Worth	0.262	15	Fort Worth	0.145	15
Houston	0.231	16	Savannah	0.106	16
Baltimore	0.178	17	Charlotte	0.012	17
Louisville	0.143	18	Houston	0.012	18
Cleveland	0.076	19	Columbus	0.011	19
Denver	0.045	20	Birmingham	0.004	20
Birmingham	0.038	21	New Orleans	<0.001	21
New Orleans	0.037	22	Columbia	<0.001	22
St. Paul	0.013	23	Memphis	<0.001	23
Columbus	0.006	24	Louisville	<0.001	24
Savannah	0.003	25	Portland	<0.001	25
Portland	0.002	26	Montgomery	<0.001	26
Charlotte	0.001	27	Denver	<0.001	27
Seattle	<0.001	28	Baltimore	<0.001	28
Memphis	<0.001	29	Omaha	<0.001	29
Spokane	<0.001	30	Pittsburgh	<0.001	30
Salt Lake City	<0.001	31	Lincoln	<0.001	31
Lincoln	<0.001	32	Chattanooga	<0.001	32
Chattanooga	<0.001	33	Spokane	<0.001	33
Montgomery	<0.001	34	Seattle	<0.001	34
Columbia	<0.001	35	Salt Lake City	<0.001	35
Wheeling	<0.001	36	Wheeling	<0.001	36
Washington	<0.001	37	Washington	<0.001	37

Note: * Indicates highest probability in the district; italics indicates actual reserve bank city with less than the highest probability in its district. Probabilities based on columns (III) and (IV) from Table 6.2

districts were first determined and then representative cities within each district were selected. If so, then basing predictions off the highest 12 probabilities would be problematic, leaving district 8 unrepresented. The asterisks next to city names in Table 6.5 indicate the highest probabilities of cities in each district. Although Pittsburgh is incorrectly predicted to be one of the twelve (albeit <50%), its probability falls below that of Cincinnati. Under the constraint of only one city per pre-determined district, then St. Louis would be predicted for district 8 (above Louisville and Memphis) and only Cincinnati over Cleveland for district 5 remains in error.

Finally, rather than comparing probabilities for cities separately in each district after the estimates are determined, we alter the specification to directly take into account the district boundaries by including a set of district dummies. Estimates then reflect differences in probability relative to other cities in the same district. Note that once district boundaries are set, the RBOC's hands are tied when it comes to districts 1, 2, 3, and 7 as there is only one applicant city for each of these districts. Including district dummies simply drops them out of the regression. This is an important consideration. If the RBOC was in essence forced to select Boston, New York, Philadelphia, and Chicago, then their individual city characteristics were irrelevant, and, for example, Philadelphia's low county support for Wilson would no longer dampen estimates of the importance the RBOC may have placed on Democratic voting for selection among cities when it had multiple cities from which to choose; conversely previous estimates may have overemphasized the importance of bank votes due to the selection of the top vote receivers Chicago, New York, and Philadelphia which may have only occurred because the RBOC had no alternative to selecting these particular cities.

Logit estimates from the Federal Reserve district-specific model are presented in the final column of Table 6.2, and associated selection probabilities appear in the second set of columns in Table 6.5. All variables now have the predicted sign and are statistically significant at better than 5%. Having more Democratic senators from its state, more Democratic House members from its state, representation on the Senator or House banking committees, greater support for the incumbent Democratic president in the most recent election, and receiving more first place votes from the banks, all made it significantly more likely the RBOC would select that city.

In addition to the pre-determined selections of Boston, New York, Philadelphia, and Chicago, the cities of St. Louis, Kansas City, and San Francisco are all predicted to be selected with nearly 100% certainty. The selection of Richmond over, in particular, Baltimore is completely predictable in this model. The lowest prediction among the top 12 belongs to Minneapolis, but this is still a 2/3rds greater likelihood of selection over St. Paul. There is exactly one city predicted to be selected in each district, and only four additional cities are estimated to have had even greater than a 1% likelihood of selection: St. Paul, Cleveland, Fort Worth, and Savannah, but none more than a 25% chance. The only error appears to be Cincinnati, which is estimated

to have a 75% chance of selection, whereas Cleveland was only given a roughly 25% chance. Pittsburgh, despite outpolling both cities, is estimated to have had a basically 0% chance of selection. Unlike the two Ohio cities of Cincinnati and Cleveland it was directly competing against for headquartering the 4th district (in addition to Columbus and Wheeling), Pittsburgh had no Democratic senators representing its state, seven fewer Democratic House Representatives from its state, and the lowest county support for Wilson (25% vote share) of all the applicant cities.[12]

The initial suggestion that the RBOC's selections were political has been supported. We have correctly predicted 11 of the 12 selected cities using only political variables. The only incorrectly predicted city selection in our preferred specification was Cincinnati over Cleveland. Although Cleveland is usually the example held up as the most clearly politically motivated choice (Miller and Genc 2002), our model suggests that the selection of Cincinnati, often suggested by critics to have been the appropriate choice, would have been more directly politically motivated. The two cities had the same state characteristics and are located in counties which had similar support levels for Wilson in the 1912 election, but banking interests preferred Cincinnati over Cleveland. The selection of Cleveland over Pittsburgh may be due to political favoritism, but its selection over Cincinnati cannot be similarly explained and may thus be the one case where the RBOC actually did not succumb to political influence, at least not from the national banks.

6.4 Conclusion

The selection of the 12 Federal Reserve headquarter cities became controversial as soon as they were announced. Although several studies have claimed politics influenced the decisions, empirical support has been either absent (McAvoy 2006) or indirect (Binder and Spindel 2013). We show that previous regression specifications relying on state level political variables are misspecified, and that supplementing a basic state political model with additional local level political variables, notably Democratic party vote share in the previous year's presidential election and national bank recorded city preference votes, results in being able to correctly predict 11 of the 12 selected cities without the need to add any "economic" variables at all. The only incorrectly predicted city selection was Cincinnati over Cleveland for the 4th district. That the original selection of the other cities can be entirely predicted on the sole basis of political variables is troubling for the founding of the Fed.

[12]With of course the exception of Washington, DC which does not cast any ballots in presidential elections.

References

Bagehot W (1873) Lombard street: a description of the money market. Henry S. King & Company, London
Beckworth D (2010) One nation under the Fed? the asymmetric effects of US monetary policy and its implications for the United States as an optimal currency area. J Macroecon 32(3):732–746
Binder S, Spindel M (2013) Monetary politics: origins of the Federal Reserve. Stud Am Political Dev 27(1):1–13
Bordo M, Roberds W (2013) The origins, history, and future of the Federal Reserve: a return to Jekyll Island. Cambridge University Press, Cambridge
Broesamle JJ (1973) William Gibbs McAdoo: a passion for change, 1863–1917. Kennikat Press, Port Washington
Brown DM (1930) Review: bankers' balances by Leonard Watkins. J Polit Econ 38(4):493–495
Chandler LV (1958) Benjamin strong: central banker. Brookings Institution, Washington, DC
Elliott MC (1914) Report of Secretary. RBOC (Records of the Federal Reserve System, Record Group 82). Technical report, National Archives II, College Park, MD
Federal Reserve Board (1915) The redistricting decision. Fed Reserve Bull 1(2):85–90
Federal Reserve Board (1959) All-bank statistics. Board of Governors of the Federal Reserve System, Washington, DC
Gerena C (2007) Federal Reserve: a division of power. Fed Reserve Bank Richmond Reg Focus 11(2):2–7
Gildea JA (1992) The regional representation of Federal Reserve bank presidents. J Money Credit Bank 24(2):215–225
Glass C (1927) An adventure in constructive finance. Doubleday, Page & Company, New York
Hammes D (2001) Locating Federal Reserve districts and headquarters cities. Region 15(3):24–27
Havrilesky T (1995) The pressures on American monetary policy. Kluwer, New York
Heckelman JC, Wood JH (2008) Joint determination of regulations by the regulator and the regulated: commercial bank reserve requirements, 1875–1979. East Econ J 34(2):158–171
Houston DF (1926) Eight years with Wilson's cabinet, 1913 to 1920. Doubleday, Page & Company, New York
Jaremski M, Wheelock DC (2017) Banker preferences, interbank connections, and the enduring structure of the Federal Reserve system. Explor Econ Hist 66(4):21–43
Kolko G (2008) Triumph of conservatism: a reinterpretation of American history, 1900–1916. Free Press, New York
Lown CS, Wood JH (2003) The determination of commercial bank reserve requirements. Rev Financ Econ 12(1):83–98
Mason DL (2004) From buildings and loans to bail-outs: a history of the American savings and loan industry, 1831–1995. Cambridge University Press, Cambridge
McAdoo WG (1931) Crowded years: the reminiscences of William G. McAdoo. Houghton Mifflin, Boston
McAvoy MR (2006) How were the Federal Reserve bank locations selected? Explor Econ Hist 43(3):505–526
McAvoy MR (2012) Bankers preferences and locating Federal Reserve bank locations. Essays Econ Bus Hist 22(1):143–169
McKinley V (2012) Financing failure: a century of bailouts. Independent Institute, Oakland
Miller JR, Genc I (2002) A regional analysis of Federal Reserve districts. Ann Reg Sci 36(1):123–138
Odell KA, Weiman DF (1998) Metropolitan development, regional financial centers, and the founding of the Fed in the lower South. J Econ Hist 58(1):103–125
Reed HL (1922) The development of Federal Reserve policy. Houghton Mifflin, Boston
Reserve Bank Organization Committee (1914a) Hearings on Federal Reserve district divisions and location of Federal Reserve banks and head offices: January 5–February 17. Government Printing Office, Washington, DC

Reserve Bank Organization Committee (1914b) Decision of the Reserve Bank Organization Committee determining the Federal Reserve districts and the location of Federal Reserve banks under Federal Reserve Act: April 2. Government Printing Office, Washington, DC
Reserve Bank Organization Committee (1914c) First-choice vote for reserve-bank cities. House Doc. 1134, 63rd Congress, 2nd session
Rockoff H (2003) How long did it take the United States to become an optimal currency area. In: Capie FH, Wood GE (eds) Monetary unions: theory, history, public choice. Routledge, New York, pp 76–103
Romer CD, Romer DH (2013) The most dangerous idea in Federal Reserve history: monetary policy doesn't matter. Am Econ Rev 103(3):55–60
Schwartz AJ (1992) The misuse of the Fed's discount window. Fed Reserve Bank St Louis Rev 74(5):58–69
Selgin G, Lastrapes WD, White LH (2012) Has the Fed been a failure? J Macroecon 34(3):569–596
Sprague OMW (1910) History of crises under the national banking system. US Government Printing Office, Washington, DC
Thornton H (1802) An enquiry into the nature and effects of the paper credit of Great Britain. Hatchard, London
Timberlake RH (1978) The origins of central banking in the United States. Harvard University Press, Cambridge
US Congress (1913) Congressional Record. Government Printing Office, Washington, DC
US Congress (1914) Congressional Record. Government Printing Office, Washington, DC
Warburg PM (1930) The Federal Reserve system: its origin and growth. Macmillan, New York
Watkins LL (1929) Bankers' balances: a study of the effects of the Federal Reserve System on banking relationships. AW Shaw Company, Chicago
Weyforth WO (1933) The Federal Reserve Board: a study of Federal Reserve structure and credit control. Johns Hopkins Press, Baltimore
White EN (2013) "To Establish A More Effective Supervision of Banking": how the birth of the Fed altered bank supervision. In: Bordo MD, Roberds W (eds) The origins, history, and future of the Federal Reserve. Cambridge University Press, Cambridge, pp 7–54
Wicker E (2005) The great debate on banking reform: Nelson Aldrich and the origins of the Fed. Ohio State University Press, Columbus
Willis HP (1923) The Federal Reserve system. Ronald Press, New York
Wood JH (2005) A history of central banking in Great Britain and the United States. Cambridge University Press, Cambridge
Wood JH (2015) Central Banking in a democracy: the Federal Reserve and its alternatives. Routledge, New York

Chapter 7
Demand for Private and State-Provided Health Insurance in the 1910s: Evidence from California

Dora L. Costa

Abstract This paper analyzes the demand for both private and state-provided health insurance in a historical context. In the case of private health insurance, I show that both health insurance and medical care were of limited use and that the relationship between income and health insurance and income and medical care was relatively weak, suggesting that money could buy little in the way of improvements in medical care. These results implied that there should be very little demand for state-provided health insurance and indeed there was not. Although the persuasiveness of interest groups such as doctors and to a lesser extend trade unions did contribute to the defeat of state-provided health insurance matter, none of the variables could explain such a resounding defeat. Evidence from newspaper editorials, advertisements, and articles suggested that the absence of consumer demand for health insurance together with concerns over the cost of state-provided health insurance defeated the measure. My findings are in contrast to those of other researchers who have emphasized the role of a politically powerful medical profession and of World War I.

7.1 Introduction

Whether health insurance should be provided by the market or by the state is a debate with a long history that in the United States first became prominent in the 1910s when social reformers turned to health insurance immediately after the passage of workers' compensation. Health insurance was viewed as actuarially and administratively the simplest branch of social insurance and the most attractive politically. Sickness affected every member of the family and sickness insurance could be adopted to the mechanism worked out for state industrial accident funds. In 1912, the year that compulsory health insurance was introduced in Britain,

D. L. Costa (✉)
University of California, Los Angeles, CA, USA
e-mail: costa@econ.ucla.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_7

the Progressive Party adopted accident, sickness, old age, and unemployment insurance as part of its platform. That same year the American Association for Labor Legislation, an organization of middle class reformers, decided to undertake an active campaign for compulsory health insurance. Commissions were created in several states to study health insurance. Health insurance bills were introduced in several state legislatures. In California there was a referendum on state-provided health insurance in the form of a proposed constitutional amendment on the 1918 ballot that would have given the legislature the right to establish a system of health insurance. All of these measures were resoundingly defeated.

The defeat of state-provided health insurance has been attributed to many causes, both general—such as the waning of Progressivism—and specific—such as opposition from doctors, insurance companies, and Christian Scientists, and the anti-German hysteria aroused by WWI (Ohio Health and Old Age Insurance Commission 1919; Anderson 1951; Viseltear 1969; Numbers 1978; Starr 1982). One aspect that has received little attention among scholars is whether there was much consumer demand for health insurance in the United States. Relatively high American incomes may have made self-insurance an option. If there was little interest in having private insurance then there should be no demand for state-provided insurance. Contemporaries were divided over whether there was consumer demand for state-provided health insurance. This paper first investigates who used private health insurance by examining consumer expenditures on private health insurance and then who demanded state-provided health insurance by examining voting on the California referendum. The two demands are intertwined. Interest group and consumer support for state-provided health insurance will depend upon satisfaction with the current system of private insurance. Conversely, pressure for national health insurance results in government policies that change private health insurance markets. Examining the demand for state-provided health insurance not only provides a quantitative assessment of consumer support for the California amendment, but also a quantitative assessment of the persuasiveness of interest groups.

The paper begins by describing the structure of private health insurance, focusing on coverage by demographic and socioeconomic characteristics, particularly income. The next section discusses the arguments made in the nineteenth century for and against health insurance and identifies the major interest groups. A quantitative analysis of the demand for state-provided health insurance is presented in the fourth section. The fifth section enhances this quantitative analysis with an examination of 1918 newspaper editorials and advertisements.

7.2 Private Health Insurance in 1918

One of the factors that a consumer considering the purchase of an insurance policy will take into account is the size of the financial risk being insured. The greater the variance confronting the consumer the higher the demand for insurance. Today the

type of coverage most commonly held is for hospital care which has the highest variance of risk. But, in the past the variance of wages lost due to illness was much greater than that due to medical costs incurred because of illness. Early medical technology provided few treatments. The coefficient of variation of lost wages was 310 and that of illness was 107.[1] Early advocates of health insurance recognized that the "insured workman . . . is very much more concerned with the size of the weekly benefit he may expect when he is compelled to 'lay off' because of ill health" than with medical benefits (Rubinow 1916). In fact, the risk that was most commonly insured against was that arising from lost wages.[2]

Wage earners in the 1910s sought to insure themselves against sickness through fraternal orders, trade union benefits, benevolent societies, commercial hospital associations, commercial insurance companies, and through working for employers who provided medical benefits.[3] Only those between age 18 to 50 and in good health, as verified by a medical exam, could join an insurance organization. However, once an individual had joined, he could continue to receive benefits even at older ages.

Sex and race were often grounds for exclusion except for insurance obtained through commercial insurance companies. Benefits could be collected after an enrollment period of 3–6 months. State social insurance commissions in California, Illinois, and Ohio estimated that about one-third of the wage earners in those states were insured. However, among prime age wage earners, the fraction was probably higher. Men joined fraternal organizations between ages 25–35 and a greater proportion of men aged 35–44 were insured relative to men aged 25–34 or younger (Emery 1993; Whaples and Buffum 1991).

Fraternal orders provided the most important form of insurance. In California an estimated 35% of all members of fraternal organizations were entitled to sick benefits. Orders paid a cash amount (from $1 to $10 per week) for a given number of weeks, generally 12 to 13, to members disabled on account of sickness.[4] The California Social Insurance Commission reported that only a minority of the lodges provided medical benefits as well as cash and that a few gave doctors' services to the immediate family of members.[5] When medical benefits were provided, they were generally through a physician employed by the fraternal organization and

[1]Calculated from California Bureau of Labor Statistics (1892) and United States Department of Labor, Bureau of Labor Statistics (1985).

[2]By the 1930s, the coefficient of variation of illness had risen and medical expenditures become the big concern.

[3]For a detailed discussion of policies see California Social Insurance Commission (1917) and Illinois Social Insurance Commission (1919).

[4]The average weekly wage of manufacturing workers was $14.97 (Inter-university Consortium for Political and Social Research 2005, Series 802–810).

[5]The use of lodge physicians varied widely. For example, in 1914 in North Adams, Massachusetts, 8000 persons out of a population of 22,000 were in the care of lodge physicians (Rosen 1977). The use of lodge physicians appears to have been more prevalent in Britain where perhaps as many as 60% of wage earners had access to lodge doctors (Naylor 1986).

paid per number of patients covered. Major operations were usually excluded from the service guaranteed as was hospital treatment. Lodge members frequently made donations to cover needs not met by insurance. Except for fraternal orders organized by ethnicity, members were generally the better paid wage earners.

Trade unions were the next largest insurance medium. Approximately 41% of union members in California and one third in Illinois were protected through their union. Membership was compulsory upon the union men who could meet the requirements. Cash benefits ranged from $1 to $10 per week and duration of benefits from 6 to 26 weeks, but on the whole the protection provided by unions was less than that offered by fraternals. Medical and hospital care were not provided.

Another form of protection was offered by the benevolent society. There were two types of benevolent societies. The first provided only cash benefits and the second owned and maintained a hospital and clinic and had a staff of visiting physicians. Members were entitled to medical, surgical, and hospital care for as long a period as necessary. Commercial hospital associations resembled this latter form of benevolent society. Members of commercial hospital associations were entitled to medical and hospital service for diseases not excluded in the contract. Diseases excluded by most contracts were venereal disease, cancer, tuberculosis, pregnancy, and pre-existing conditions.

Commercial insurance policies were relatively unimportant. In California only about one sixth as many workers were commercial policy holders as were members of fraternal organizations. These policies were written to sometimes cover sickness only, sometimes accident only, and sometimes both sickness and accident. There were two types of commercial insurance policies—commercial and industrial. Industrial policies were targeted towards wage-earning groups, generally the more highly paid wage earners such as mechanics and other skilled labor, and were offered at a monthly premium. Industrial policies insuring against disability arising from one or more of a certain number of diseases or excluding disability due to certain diseases, such as tuberculosis, were also sold. Doubtful risks were given medical examinations. Benefits under most industrial policies were limited to 6 or 8 months and many of the policies restricted the payment of the usual weekly or monthly benefit in case of certain chronic diseases and diseases of long duration to a fraction of the maximum. Commercial policies were sold primarily to business and professional men, farmers, and others of the non-wage earnings classes on an annual premium plan. They provided weekly benefits to compensate for time lost from disability due to sickness or accident, ranging from $10 to $50 or more and other benefits.

Employers would sometimes provide cash benefits in case of sickness and less often medical, surgical, or hospital care. In Illinois approximately 7.5–10% of the wage earners of the state were employed in establishments in which a benefit system was in operation. Payment into establishment funds might be either by employers only, by both employers and employees, or by employees only. Establishment funds were more common among oil, lumber, and mining companies and among railroads than among manufacturing firms, and were expanding. Beginning in 1912, larger

employers also provided group life insurance. An employee covered under a group contract who became totally and permanently disabled could collect the income payable in the event of a death as income over a several year period.[6]

The relation between health insurance and worker and household characteristics can be more formally investigated using the 1917–1919 Bureau of Labor Statistics Cost of Living Study which contains information on 12,817 families of wage earners or salaried workers in 99 cities in 42 states (United States Department of Labor, Bureau of Labor Statistics 1985). Interviewed households were restricted to those where both spouses and one or more children were present, where salaried workers did not earn more than $2000 a year, where families had resided in the same community for a year prior to the survey, where families did not take in more than three boarders, where families were not classified as either "slum" or charity, and where non-English speaking families had been in the U.S. five or more years. The survey therefore oversamples craft and other high skilled workers relative to factory operatives and laborers and professionals and semi-professionals. Interviewers may also have picked a specific neighborhood in a city.[7] Black and unknown race households were excluded from the sample, leaving 11,933 households.

The 1917–1919 Cost of Living Survey contained questions on how many people in the household had health insurance and how many people had accident insurance. These questions were answered mainly by those insuring through insurance companies, mutual associations, and employing associations. Unless the primary purpose of a lodge or union was health insurance or unless the insurance component could be separated from other lodge or union dues, insurance through a fraternal organization or through a union would be counted as expenditures on lodges or unions. Also, because health and accident insurance policies were often sold at one rate of premium, the extent of health insurance alone will be underestimated by the health insurance question. In fact, only 10% of households reported having health insurance and 19% either health or accident insurance.[8] If 35% of all fraternals and 41% of all unions provided health insurance as the California Social Insurance Commission estimated then about 33–42% of all households surveyed had health insurance.[9]

A household is assumed to insure if the difference between the risk premium that they would be willing to pay against a risky event and the amount the insurance company charges for risk bearing is greater than zero. The premium that the household will be willing to pay will depend upon income and upon age, because perceptions of risk may change with age, and upon lodge and union membership because these may provide alternative forms of insurance. The amount charged for

[6]For more details on group life insurance see Hoffman (1917).

[7]There are also large numbers of workers from the same industry or firm close together in the sample.

[8]The numbers are the same for California alone.

[9]In the data only 10% of households reporting lodge expenditures also reported having health insurance. Similarly, only 10% reporting union expenditures also reported having health insurance.

risk bearing is likely to be related to age, lodge and union membership, and because of differences by industry in the prevalence of employer establishment funds and occupational hazards, industry.

The logarithm of husband's yearly earnings, wife's yearly earnings, children's yearly earnings, income from rent and interest, income from boarders, income from gardens, and all other income is used in the estimation equations.[10] State social insurance reports regarded income as an important predictor of insurance. Among Illinois husband and wife wage-earner households 38% of high income husbands were insured but only 24% of low income husbands. The number of non-working children is also included in the estimation equations. Age is entered quadratically.[11] Dummies are used for lodge and union membership. Occupations were divided into five classes—(1) professionals, semi-professionals, and proprietors, (2) clerical and sales, (3) craft and skilled, (4) service, including personal, and protection, and (5) unskilled laborers. Industries were divided into 14 categories. Region of residence was divided into eight regions.[12] Probit equations of the probability of holding health insurance and either health or accident insurance are given in Table 7.1.

The probability of holding insurance rose with age at a decreasing rate. Both lodge and union members were less likely to have insurance, probably because many of them already had insurance through their organizations and did not answer the health insurance question affirmatively. There were large differences by industry in the probability of holding insurance. Workers in extractive industries were the most likely to have insurance. Railroad employees were the next most likely but did not approach the probability of employees in extractive industries. Husband's earnings were the most important component of total earnings in determining the probability of insurance and the amount spent on insurance. However, the impact of earnings was small. The elasticity of insuring with respect to husband's earnings was only 0.10 for health insurance alone and 0.23 for health or accident insurance. When industry dummies were excluded the elasticity for health insurance rose to 0.20.[13]

The relationship between income and the probability of insurance might be weak because of bias in the dependent variable. Many lodge and union members were not listed as being insured and these workers tended to be relatively well paid.

[10]Ten cents was added to all incomes to avoid taking logarithms of zero. The use of dummy variable categories rather than the logarithm indicated that the probability of insurance rose with income but at a decreasing rate. Income was divided into its various components because health insurance made up for the husband's lost earnings, but the results remain unchanged if all components of income are added together.

[11]Tests indicated that this was the best specification for age.

[12]Trying to add state fixed effects to the equation led to multicollinearity problems.

[13]The low elasticity with respect to husband's earnings contrasts with Whaples and Buffum (1991), who find that among Michigan furniture workers in 1889 a 10% rise in wages increased the probability of purchasing sickness/accident insurance by about 10%. The sample used by Whaples and Buffum (1991) may have consisted of individuals less able insure through their families, since it contained unmarried individuals, a large proportion of immigrants, and a large fraction of urban households.

Table 7.1 Probit estimates of determinants of holding health insurance and either health or accident insurance, 1917–1919

11993 obs		Health insurance Pseudo $R^2 = 0.08$			Health or accident Pseudo $R^2 = 0.10$		
Variables	Mean	Coef	Std err	$\frac{\partial P}{\partial x}$	Coef	Std err	$\frac{\partial P}{\partial x}$
Intercept		−3.6342‡	0.5594		−3.7208‡	0.4347	
Log of yearly							
Husband's earnings	7.17	0.1314*	0.0685	0.0210	0.3011‡	0.0525	0.0589
Wife's earnings	−1.72	0.0024	0.0012	0.0004	0.0084	0.0074	0.0016
Children's earnings	12.0	0.0162†	0.0074	0.0026	0.0069	0.0061	0.0014
Rent and interest	−1.71	0.0213†	0.0089	0.0034	0.0189‡	0.0076	0.0037
Income from boarders	−1.97	0.0050	0.0113	0.0008	0.0044	0.0096	0.0008
Income from gardens	−0.13	0.0003	0.0068	0.0000	0.0101*	0.0057	0.0020
Other income	−1.62	0.1284‡	0.0077	0.0205	0.1146‡	0.0071	0.0224
Number non-working children	2.27	−0.0038	0.0178	−0.0006	−0.0233	0.0153	−0.0046
Age	36.93	0.0413‡	0.0142	0.0066	0.0221†	0.0112	0.0043
Age squared		−0.0004†	0.0002	−0.0000	−0.0002*	0.0001	−0.0000
Dummy=1 if member							
Lodge	0.29	−0.0397	0.0368	−0.0064	0.0177	0.0308	0.0035
Union	0.31	−0.0524	0.0376	−0.0084	−0.0467	0.0318	−0.0091
Dummy=1 if occupation							
Professional or semi-professional	0.07	–			–		
Clerical or sales	0.15	0.0680	0.0816	0.0109	−0.0037	0.0651	−0.0007
Skilled	0.38	0.0565	0.0736	0.0090	−0.0281	0.0592	−0.0055
Service	0.06	0.0357	0.0989	0.0057	−0.0553	0.0809	−0.0108
Laborer	0.34	0.0875	0.0754	0.0140	−0.0317	0.0608	−0.0062
Dummy=1 if industry							
Personal, repair, service, sales	0.12	–			–		
Government	0.07	0.3903‡	0.0820	0.0625	0.3555‡	0.0665	0.0695
Construction	0.05	−0.0438	0.1130	−0.0070	−0.1898†	0.0920	−0.0371
Steel or steel products, chemical, rubber	0.16	0.3623‡	0.0731	0.0581	0.2727‡	0.0584	0.0533
Railroad	0.11	0.4720‡	0.0762	0.0755	0.5218‡	0.0604	0.1020
Other transport and transport production	0.13	0.2712‡	0.0761	0.0434	0.1878‡	0.0607	0.0367

(continued)

Table 7.1 (continued)

11993 obs		Health insurance Pseudo $R^2 = 0.08$			Health or accident Pseudo $R^2 = 0.10$		
Variables	Mean	Coef	Std err	$\frac{\partial P}{\partial x}$	Coef	Std err	$\frac{\partial P}{\partial x}$
Lumber	0.04	0.2733‡	0.1043	0.0437	0.3238‡	0.0841	0.0633
Utilities	0.02	0.2134	0.1404	0.0341	0.1608	0.1136	0.0314
Food manufacturing	0.06	0.2079‡	0.0933	0.0333	0.0913	0.0753	0.0179
Printing	0.02	0.2488*	0.1363	0.0398	0.1060	0.1139	0.0208
Textiles	0.05	0.3800‡	0.0946	0.0608	0.2129‡	0.0800	0.0416
Extractive industries	0.04	0.8065‡	0.0920	0.1290	0.8485‡	0.0789	0.1659
Communications and financial	0.02	0.0090	0.1376	0.0014	−0.0604	0.1063	−0.0118
Miscellaneous manufacturing or unknown	0.10	0.2124‡	0.0803	0.0340	0.1040	0.0649	0.0203

Note: Regional dummies were included in both specifications. The symbols ∗, †, and ‡ indicate that the coefficient is significantly different from 0 at least the 10, 5, and 1% level, respectively. $\frac{\partial P}{\partial x} = \beta \frac{1}{n} \sum \phi(x'\beta)$, where ϕ is the standard normal density, and is in probability units

If it is assumed that all lodge and union members had health insurance then the elasticity of insuring with respect to husband's earnings is still only 0.18. A more likely explanation for the weak relationship between income and the probability of insurance is that higher income households could depend upon savings and lower income households upon either increased labor participation of other family members, reductions in expenditures, or charity. Charity was provided by county governments and by private organizations. Free medical, dental, and nursing care could be obtained from county hospitals, city governments, neighborhood health centers, or free private clinics. California made its counties responsible for the health care of poor persons in 1855 (Shonick 1995). In 1920 the United States spent 0.40% of its GDP on government subsidies for health care, increasing spending from 0.26% in 1910 and spending slightly less than Britain (0.56%) but more than Sweden or France (0.25% each) (Lindert 1994). Friends and neighbors were other important sources of non-market insurance (Whaples and Buffum 1991).

The opponents of state-provided health insurance argued that the reason so few wage-earners were insured was because there was simply not that much need for health insurance. Sickness was infrequent and easily covered by savings. The California Social Insurance Commission (1917) estimated that wage earners lost at most 6 days per year or about one work week due to illness. Those who experienced some sickness lost on average 5 weeks. In the 1917–1919 Cost of Living Study (United States Department of Labor, Bureau of Labor Statistics 1985), the average household had yearly savings worth 2.8 weeks' of the husband's weekly wage. The average cost of health insurance to those households that did have it was equal to one-half of the husband's weekly earnings. When households were divided into quartiles according to husband's weekly earnings those in the lowest quartile had yearly savings totalling 1.2 weeks of the husband's weekly earnings and those in the highest quartile yearly savings totalling 4.4 weeks. Insurance through a fraternal was

less expensive. The California Social Insurance Commission (1917) estimated that many fraternals charged $9 per year, or 35% of the weekly earnings of husbands in the 1917–1919 Cost of Living Survey (United States Department of Labor, Bureau of Labor Statistics 1985), for benefits of $5–$10 per week for up to 13 weeks. Thus, although the benefits were less generous than those of commercial insurance companies, premiums were less expensive and benefits more than covered expected sickness days.

Proponents of state-provided health insurance acknowledged that fraternals could provide insurance coverage at a fraction of commercial insurers' costs because local lodges had more complete information about their members than insurance companies ever could, but questioned whether relatively small lodges could be safe sources of insurance. They admitted that commercial insurance companies were financially stable, but argued that the premiums of commercial insurance companies were high because of the moral hazard problem. The failure of individuals to insure could therefore be attributed to the high costs of insurance. Recent work, however, suggests that fraternal lodges had almost no probability of being bankrupted by high claims because surplus revenues were invested in assets which generated income that subsidized lodge operations and benefit payments (Emery 1993). Access to health insurance thus appears not to have been determined primarily by income.

The relatively weak relationship between health insurance and income is mirrored by the relatively weak relationship between medical expenditures and income. The share of medical expenditures is constant across income categories (see Table 7.2).

Table 7.2 Budget share devoted to medical care and share of medical care budget devoted to various medical expenditures by annual income

	All households	Household income				
		< $1176	$1176–1350	$1350–1550	$1550–1827	≥ $1827
Direct medical expenditures	4.1	4.0	4.1	4.2	4.3	4.0
Medical, health and accident						
Insurance expenditures	4.2	4.2	4.3	4.4	4.6	4.3
As % direct medical expenditures						
Doctors	47.8	49.9	49.8	47.7	46.7	45.1
Drugs	25.4	29.4	25.7	24.5	24.2	23.4
Dentists	15.1	10.7	13.7	15.1	16.9	19.2
Nurses	3.4	3.3	3.5	3.7	3.7	2.9
Eye glasses	4.3	3.4	3.6	3.7	3.7	5.6
Hospitals	0.1	0.1	0.1	0.1	0.1	0.1
Other medical products	0.3	0.3	0.4	0.3	0.3	0.2

Calculated from United States Department of Labor, Bureau of Labor Statistics (1985) restricted to white households

How the medical care budget was spent was also relatively similar across income categories. The greatest expenditures, in order of magnitude, were on doctors, drugs, and dentists. Although the wealthier spent relatively more on dentists and eye glasses and relatively less on doctors and drugs, income does not appear to determine access to medical care.[14] Medical knowledge was such that there was little that money could buy, suggesting that even if health insurance had paid for both lost wages and for medical expenditures there would still be little demand for it.

7.3 Debating State-Provided Health Insurance

The health insurance plan that received the most publicity was the American Association for Labor Legislation's (AALL) "Standard Bill" published in 1915. The plan was similar to that prevailing in Germany. Health insurance was to be compulsory for every employed person earning $1200 per year or less, with special provisions covering casual and home workers and those who wished to insure voluntarily under the act.[15] Contributions for benefits were to come from the state, employees, and from employers. On the fourth day of illness, a cash benefit equal to two-thirds of the weekly wage was to be paid for a period not exceeding 26 weeks in a year. All necessary medical, surgical, and nursing attendance and treatment were to be furnished from the first day of illness for a period not to exceed 26 weeks in any one year.[16] Provisions were also made for maternity and funeral benefits. Medical service was to be provided by insurance carriers through either (1) a state panel of physicians from which patients could choose their doctor; (2) salaried physicians employed by the carrier among whom the insured had reasonable free choice; and (3) district medical officers, engaged for treatment of insured persons in prescribed areas. These insurance carriers might be either fraternals, trade funds, establishment funds, trade unions, or state local funds. However, the 40% employer contribution went to the state fund alone, thus effectively excluding all other insurance carriers. Commercial insurance companies were explicitly excluded. The state local funds were to be centrally administered by a State Social Health Insurance Commission within a multi-tiered bureaucratic organization. Renumeration of physicians was to be set at a rate approved by the Commission.

[14]There was no evidence that those living in smaller cities had less access to medical care. In fact, expenditures on medical care were somewhat greater in smaller cities.

[15]The sum of $1200 was above the average earnings of all wage earners.

[16]Hospital care was simply not a concern in the 1910s. As seen in Table 7.2 only a small share of all expenditures on medical care went to hospital care. County and private charity expenses were relatively small as well. Only 13% of all expenditures on medical care and health insurance in consumer income accounts went to hospital expenditures (calculated from Dewhurst and Associates (1955)).

Although the California Social Insurance Committee did not explicitly draft a health insurance bill, in their final report the Committee proposed a health insurance scheme slightly different from the AALL's. Insurance would be compulsory for all workmen below a specified income level, perhaps $1200 or $1500. Under their proposed plan a state health insurance fund would be established. This fund would be the sole carrier of medical benefits and one but not the sole carrier of cash benefits. Workmen would pay the entire cost of the cash benefit and would choose their own insurance carriers.[17] Thus, unlike the AALL plan, unions, fraternal societies, and other voluntary organizations would not be excluded.[18] Employers and the state would contribute to the medical benefit. Physicians would be organized in districts and supervised by medical inspectors. Their fees would be set. The insured would be able to choose from any physician registered with the district. There would be limits on the duration of both medical and cash benefits. The Commission favored a somewhat different organizational form than that proposed by the AALL. Because the Social Insurance Commission believed that there were constitutional obstacles in the path of legislative action, they recommended that a constitutional amendment enabling the legislature to establish a system of health insurance be submitted to the people (California Social Insurance Commission 1917).[19]

The arguments for and against health insurance were publicized on both the national and the state level.[20] Proponents of health insurance argued that the state could provide larger and more extensive benefits than private insurance carriers because by insuring a larger group of people aggregate risk would be lower and because many selling and administrative costs would be eliminated. Furthermore, the results of sickness were so disastrous that sickness was a community concern. Proponents of health insurance stressed that ill health often led to families becoming charity cases. Another argument made by the proponents of health insurance was that only by compulsion was it possible to distribute throughout industry the burden of sickness. Sickness was a result of the action or inaction of the state (public health measures), of employers (working conditions), and of employees (personal health

[17]The Commission believed that by having workmen pay the entire cost of the cash benefit, malingering could be avoided.

[18]In its inclusion of fraternal societies, the California Social Insurance Commission's plan more closely resembled the British rather than the German plan.

[19]The proposal of an enabling amendment should not be regarded as unusual. In California workers' compensation was initially struck down by the courts and passed only by amending the California constitution.

[20]See Ohio Health and Old Age Insurance Commission (1919), Illinois Social Insurance Commission (1919), California Social Insurance Commission (1917), Hoffman (1917), Warren and Sydenstricker (1916), among others.

habits). Because the costs of insurance were spread among all three groups, all three groups would take actions to improve health and to lower premiums.[21] The average health of the nation would thus improve.

The opponents of national health insurance argued that state provided health insurance would not improve health. Public health measures were needed. They pointed out that in the German case there was no integration of health insurance and public health agencies and that the health of the United States compared favorably with that of Germany. They also argued that poverty would not be avoided by compulsory health insurance. The unemployed were the ones likely to be left out of the system. The poor would be better helped by a state medical service, better poor relief, and the establishment and adequate maintenance of community hospitals. One of the big concerns of opponents was cost. The California Social Insurance Commission (1917) estimated that costs would range from $14,651,000 to $28,780,000 per annum.[22] There was fear that costs would be raised by malingering on the part of individuals and by graft and mismanagement and that only "politicians and job chasers" would benefit.[23] In addition to the total costs, there was concern with the distribution of costs. Why should all of society have to pay for sickness caused by personal habits or by industry specific working conditions? Shouldn't industry specific illness be covered by workers' compensation instead?

The major interest groups lobbying for or against state provided health insurance were middle class Progressive organizations such as the AALL, unions, employers, doctors, druggists, and insurance companies and fraternals.[24] Unions were divided. The American Federation of Labor, led by Samuel Gompers, was opposed. In California the San Francisco Labor Council was opposed, while the California Federation of Labor and the California Building Trades Council were in favor (Skocpol 1992, pp. 238, 244). Opposition to the measure arose from the belief that the measure might undermine union activity, prove to be a palliative and a substitute for better wages, hours, and conditions of labor, create class divisions, and induce further state regulation of the worker's purely personal affairs.

[21] Oddly enough, the issue of free riding never came up.

[22] There was a high variance in estimates of the cost of insurance. The AALL estimated that 4% of payroll would be needed to cover lost wages, medical aid, and maternity and funeral benefits. The Illinois Social Insurance Commission (1919) estimated that 7.5% of payroll would be needed to cover lost wages and medical care alone (Starr 1982).

[23] This was argued in one of the pamphlets sponsored by the California Research Society of Social Economics. In May of 1917 there was a full page advertisement in the annual edition of the San Francisco Chronicle paid for by the Insurance Federation of California in which it was stated that health insurance would become political graft so gigantic that the political party in power at the time health insurance was adopted could perpetuate itself (Viseltear 1969).

[24] See Ohio Health and Old Age Insurance Commission (1919) and Starr (1982).

Employers argued that the initial expense of the insurance could not be absorbed in the costs of production and shifted to consumers. They feared that employers in states which adopted it would be at a disadvantage in competition with those of other states. The National Association of Manufactures took no stand on health insurance, but the National Industrial Conference Board was represented in various state legislative hearings and at various public forums in California.

Although the American Medical Association (AMA) had established a social insurance commission in 1915 and although many of the most prominent supporters of the AALL's plan were physicians, the AMA took no stand on health insurance, except to insist that any proposed legislation should provide for freedom of choice of physician by the patient; payment of the physician in proportion to the amount of work done; the separation of the function of medical official supervision from the function of daily care of the sick; and the adequate representation of the medical profession on the appropriate administrative bodies. Some doctors argued that health insurance would benefit the medical profession by increasing the size of the average physicians' practice. Table 7.2 suggests that any increases would have been small. Nonetheless, some of the leaders of the California state medical society favored compulsory health insurance. The state medical society took no stand, but a group of doctors formed an independent association to oppose the constitutional amendment. Doctors' opposition sprang in part from the fact that no specific bill had been drafted. Before the California popular vote doctors sent open letters and pamphlets to their patients and other physicians warning them against health insurance (Viseltear 1969).

Other medical personnel, such as nurses, dentists, druggists, occultists, and hospitals demanded that they too be represented in any state health insurance bureaucracy, fearing that doctors might benefit at their expense.[25] Insurance companies and fraternals also feared being shut out from health insurance and under the AALL's plan they would have been. They published pamphlets, paid for newspaper advertising, and when the proposed constitutional amendment was placed on the California ballot, sent personal communications to their members, asking that they vote against the amendment. Contemporaries estimated that they reached over 200,000 voters (Ohio Health and Old Age Insurance Commission 1919).

Another group whom contemporaries believed to be a potent force in the battle over health insurance were Christian Scientists. Christian Scientists wrote editorials and organized "publicity committees" in San Francisco and Los Angeles to assist them in their campaign against health insurance.

[25]Druggist could point to the "floating six-pence" of the British health insurance act, whereby the physicians' fund benefited from a reduction in the amount of drugs prescribed.

7.4 Voting on State-Provided Health Insurance

Both the State Senate and Assembly passed the proposed constitutional amendment in 1917. In the Senate the vote was 32 to 3 with the only nay votes among senators from Los Angeles.[26] In the Assembly the vote was 55 to 11. The only predictors (positive) of a nay vote in the Assembly were whether a county was either a big lumber or mineral producer, but even in these counties there was little opposition. After the constitutional amendment was passed by the Legislature, the campaign against state-provided health insurance intensified. On election day, the health insurance amendment was defeated by 358,324 votes to 133,858.

Although 714,525 voters showed up at the polls on election day, only 492,182 voted on the health insurance amendment. In contrast, 688,670 voted for a gubernatorial candidate and 582,131 voted on the prohibition amendment. Because so many voters abstained on the health insurance amendment, I use the voting model of Deacon and Shapiro (1975). In their model, an individual is assumed to vote for health insurance if utility under the health insurance outcome is perceptibly greater than utility under the no health insurance outcome. If the difference in utility under the two alternative outcomes is not perceptible, the individual will be indifferent. Suppose that the difference between utilities resulting from the passage and the defeat of health insurance is Δu_i for individual i. An individual will vote yes if $\Delta u_i > \delta_i$, no if $\Delta u_i < -\delta_i$, and abstain otherwise. The term δ_i can be interpreted as a measure of the precision with which and individual is able to distinguish between the two alternatives. I assume that $\delta_i = x_i\gamma'$, where x_i is a vector containing variables that reflect experience, predisposition, and contextual knowledge. I assume that Δu_i is distributed logistically with mean μ_i, that is $P(\Delta u_i < x_i) = \Lambda(x_i - \mu_i)$, where Λ is the logistic distribution. I assume that $\mu_i = z_i\alpha' + x_i\beta'$, where z_i is a vector containing variables of socioeconomic characteristics and voter attributes that are relevant to the health insurance referendum, such as the proportion of voters in an affected group.

The model is estimated with county level data using Berkson's minimum chi-squared technique by jointly estimating

$$\begin{aligned} \ell(Y) &= z_i\alpha' + x_i(\beta' - \gamma') \\ \ell(Y) + \ell(N) &= -2x_i\gamma' \end{aligned} \tag{7.1}$$

where $\ell(Y) = \log(\frac{P(Y)}{1-P(Y)})$, $\ell(N) = \log(\frac{P(N)}{(1-P(N)})$, $P(Y)$ is the probability of voting yes, and $P(N)$ is the probability of voting no (Deacon and Shapiro 1975). The estimated coefficients of the first equation give the overall effect of the independent variables upon the probability of voting in favor of the proposition, while the

[26] By 1916 southern California was the stronghold of anti-Progressive sentiment. Progressive strength was strongest among San Francisco workers and the Catholic foreign-stock counties of the Bay Area (Rogin and Shover 1970).

estimated coefficients of the second equation give the overall effects of voting on the initiative at all. The values of α and β can be derived from the estimated equations and give the effects of the independent variables upon expected changes in utility. The estimates of γ give the impact of the independent variables upon the perception threshold.

Voter attributes that are directly relevant to the health insurance referendum are, per voting age individual, the value of crops, of manufactures, and of mineral wealth, and the taxable wealth of railroads; whether a county was a lumber producer; city size; and the number of hospital beds per capita. Demand for health insurance should vary by type of industry. As previously noted, workers in the railroad, lumber, and mining industries were already more likely to have insurance than workers in other industries. If they were satisfied with their existing benefits they might not wish to try a new system. Farm communities in which self-employment was high probably saw the measure as benefiting largely manufacturing. Workers in manufacturing might favor the health insurance amendment because they would receive the benefit but part of the cost would be paid from general taxation. Areas without medical facilities might regard access rather than payment as one of the true problems.

Variables that reflect experience, predisposition and contextual knowledge are the logarithm of taxable wealth per voting age individual; for the voting age population, the percentage of foreign-born, illiterates, union members, and doctors; the fraction of members of religious denominations who were Christian Scientists; and the percentage of the vote cast for Wilson and Benson in the 1916 presidential election. Although numerically small, doctors, Christian Scientists, and union members are alleged to have been effective advocates for or against health insurance. Supporters of Benson (the Socialist candidate) and of Wilson (the Democratic candidate) in the 1916 presidential election should be more likely to favor insurance compared to supporters for Hughes (the Republican candidate). In California, workers and Progressives voted for Wilson (Rogin and Shover 1970).

Two specifications are given (see Tables 7.3 and 7.4). The first specification contains the past voting variables. The second does not. The estimated coefficients in the first column of Table 7.3 thus indicate the residual effects of socio-economic and demographic characteristics on the probability of voting for the amendment and the residual effects of political affiliation controlling for socioeconomic and demographic characteristics on the probability of voting for the amendment. But, if political affiliation is a function of the given socioeconomic and demographic characteristics, then the full impact of the socioeconomic and demographic characteristics is indicated by the second specification. However, if any socioeconomic and demographic variables that determine party affiliation have been incorrectly excluded from the specification, then the first equation is the correct specification.[27]

[27]The problem then becomes one of interpreting past voting behavior. Peltzman (1984) finds that with increasingly better economic variables the impact of party affiliation declines and suggests that party affiliation reflects not ideology but economic self-interest. Poole and Rosenthal (1993) find that once an ideology measure has been used, the marginal explanatory power of the economic self-interest variables is minimal and hence argue that voting is best described by ideology.

The regression results indicate that industry, which predicted the probability of having health insurance, was negatively, but insignificantly, related to the probability of voting for the constitutional amendment. City size was positively related to the probability of voting for the amendment, and significantly so when past voting variables were omitted from the regression. When the percentage of voters over age 45 and of women of voting age was added to the regressions, the coefficients were positive but highly insignificant, suggesting that the unavailability of health insurance at older ages or to women did not lead to age or sex differences in voting patterns. Wealth, another predictor of the probability of having health insurance was in Table 7.3 a significant, negative predictor of the probability of voting for

Table 7.3 Determinants of vote on 1918 health insurance referendum

58 obs, adjusted $R^2 = 0.95$		$\ell(Y)$		$\ell(Y) + \ell(N)$		μ	δ
County level variables	Mean	α, $(\beta - \gamma)$	$\frac{\partial L}{\partial x}$	-2γ	$\frac{\partial L}{\partial x}$	α, β	γ
Intercept		−0.4850		−0.3894		−0.4850	0.1947
		(1.0621)		(0.9440)		(1.0621)	(0.4720)
Value per voting age individual							
Logarithm of taxable wealth	8.08	−0.2510†	−0.0101	−0.1588	−0.0064	−0.1716	0.0794
		(0.1276)		(0.1122)		(0.2167)	(0.0561)
% voting age pop doctors	0.21	−1.8259‡	−0.0731	−0.4400	−0.0176	−1.6059	0.2200
		(0.6676)		(0.5464)		(1.0822)	(0.2730)
% religious members Christian Scientists × 100	14.34	0.0010	0.0000	0.0003	0.0000	0.0008	−0.0002
		(0.0051)		(0.0038)		(0.0079)	(0.0019)
% voting age pop union members	3.40	0.0154	0.0006	0.0113	0.0005	0.0098	−0.0056
		(0.0120)		(0.0118)		(0.0219)	(0.0059)
% voting age pop illiterate	6.12	−0.0284	−0.0011	−0.0464‡	−0.0019	−0.0053	0.0232‡
		(0.0207)		(0.0180)		(0.0350)	(0.0090)
% voting age pop foreign-born	24.12	0.0100	0.0004	0.0269‡	0.0011	−0.0035	−0.0315‡
		(0.0063)		(0.0054)		(0.0105)	(0.0027)
% votes cast for Wilson in 1916	49.73	0.0256‡	0.0010	−.0010	−0.0000	0.0261*	0.0005
		(0.0097)		(0.0079)		(0.0158)	(0.0040)
% votes cast for Benson in 1916	5.22	0.0003	0.0000	−0.0535†	−0.0021	0.0272	0.0268†
		(0.0249)		(0.0229)		(0.0440)	(0.0115)
Value per voting age individual							
Crops	484.54	−0.0001	−0.0000	–		−0.0001	–
		(0.0001)		–		(0.0001)	–
Manufacturing products	0.65	0.0187	0.0008	–		0.0187	–
		(0.0246)		–		(0.0246)	–
Mineral products	45.32	−0.0003	−0.0000	–		−0.0003	–
		(0.0004)		–		(0.0004)	–

(continued)

Table 7.3 (continued)

58 obs, adjusted $R^2 = 0.95$		$\ell(Y)$		$\ell(Y) + \ell(N)$		μ	δ
County level variables	Mean	$\alpha, (\beta - \gamma)$	$\frac{\partial L}{\partial x}$	-2γ	$\frac{\partial L}{\partial x}$	α, β	γ
Railroad assessments	215.81	−0.0001 (0.0001)	−0.0000	– –		−0.00001 (0.0001)	– –
Dummy=1 if county lumber producer	0.17	−0.0951 (0.0822)	−0.0038	– –		−0.0951 (0.0822)	– –
Dummy=1 if city of 25,000+	0.12	0.0861 (0.0743)	0.0034	– –		0.0861 (0.0743)	– –
No of hospital beds per capita	0.01	−0.1766 (1.7121)	−0.0071	– –		−0.1766 (1.7121)	– –

Note: Standard errors are in parentheses. The symbols ∗, † and ‡ indicate that the coefficient is significantly different from 0 at least the 10, 5, and 1% level, respectively. $\frac{\partial L}{\partial x} = \beta \frac{1}{n} \sum L(1 - L)$, where β is the vector of estimated coefficients, L is the logistic distribution, and $\frac{\partial L}{\partial x}$ is in probability units. Sources are as follows. Voting data is from California Registrar of Voters (1918) and Clubb et al. (2006). Assessed values are from California State Board of Equalization (1918). Mineral production values are from California Secretary of State (1913). Values are for 1910. Information on lumber production within a county is from California State Tax Commission (1917). Union membership was estimated from California Bureau of Labor Statistics (1918). Religious membership was extrapolated from United States Bureau of the Census (1906, 1926). The number of hospital beds and of doctors by county was compiled from American Medical Association (1918). All other information is from Inter-university Consortium for Political and Social Research (2005)

health insurance, an insignificant, negative predictor of abstention and of the mean difference between utilities resulting from the act's passage and its defeat, μ. When past voting is omitted from the regression equations, the coefficient on wealth becomes insignificant.

The regression results also show that interest groups were important. The fraction of doctors within a county was a significant, negative predictor of the probability of voting for the constitutional amendment and a significant, negative predictor of mean expected changes in utility when past voting was omitted from the regression. Voters in areas where there was a high proportion of doctors clearly did not approve of the amendment. The fraction of Christian Scientists becomes a significant predictor of both the probability of voting for the measure and expected changes in utilities only when voting variables are omitted from the regression, perhaps because of the strong negative correlation ($\rho = -0.52$) between percent Christian Scientist and percent voting for Wilson. The fraction of voters who were union members was positively associated with the probability of voting for the constitutional amendment. When past voting behavior was omitted from the regression, the association was significant. However, the fraction of union members was not a significant predictor of differences in utility.

Table 7.4 Determinants of vote on 1918 health insurance referendum, excluding political affiliation

58 obs, adjusted $R^2 = 0.94$		$\ell(Y)$		$\ell(Y) + \ell(N)$		μ	δ
County level variables	Mean	$\alpha, (\beta - \gamma)$	$\frac{\partial L}{\partial x}$	-2γ	$\frac{\partial L}{\partial x}$	α, β	γ
Intercept		0.3866		−0.9615		0.3866	0.4808
		(1.0387)		(0.9446)		(1.0387)	(0.4723)
Value per voting age individual							
Logarithm of taxable wealth	8.08	−0.1340	−0.0056	−0.1328	−0.0055	−0.0676	0.0664
		(0.1320)		(0.1152)		(0.2233)	(0.0576)
% voting age pop doctors	0.21	−2.1387‡	−0.0891	0.0030	0.0001	−2.1402†	−0.0015
		(0.6821)		(0.5468)		(1.0932)	(0.2734)
% religious members Christian Scientists × 100	14.34	−0.0106‡	−0.0004	0.0026	0.0001	−0.0119†	−0.0013
		(0.0030)		(0.0026)		(0.0051)	(0.0013)
% voting age pop union members	3.40	0.0237†	0.0010	0.0187	0.0008	0.0143	−0.0094
		(0.0121)		(0.0117)		(0.0215)	(0.0059)
% voting age pop illiterate	6.12	−0.0432†	−0.0018	−0.0392†	−0.0016	−0.0236	0.0196†
		(0.0203)		(0.0168)		(0.0332)	(0.0084)
% voting age pop foreign-born	24.12	0.0030	0.0001	0.0215‡	0.0009	−0.0078	−0.0108‡
		(0.0063)		(0.0052)		(0.0101)	(0.0026)
Value per voting age individual							
Crops	484.54	−0.0002	−0.0000	–		−0.0002	–
		(0.0001)		–		(0.0001)	–
Manufacturing products	0.65	0.0120	0.0005	–		0.0120	–
		(0.0267)		–		(0.0267)	–
Mineral products	45.32	0.0000	0.0000	–		0.0000	–
		(0.0005)		–		(0.0005)	–
Railroad assessments	215.81	−0.0001	−0.0000	–		−0.0001	–
		(0.0001)		–		(0.0001)	–
Dummy=1 if county lumber producer	0.17	−0.1326	−0.0055	–		−0.1326	–
		(0.0871)		–		(0.0871)	–
Dummy=1 if city of 25,000+	0.12	0.1701†	0.0071	–		0.1701†	–
		(0.0742)		–		(0.0742)	–
No of hospital beds per capita	0.01	−1.7573	−0.0732	–		−1.7573	–
		(1.8047)		–		(1.8047)	–

Note: Standard errors are in parentheses. The symbols † and ‡ indicate that the coefficient is significantly different from 0 at least the 5% and 1% level, respectively. $\frac{\partial L}{\partial x} = \beta \frac{1}{n} \sum L(1 - L)$, where β is the vector of estimated coefficients, L is the logistic distribution, and $\frac{\partial L}{\partial x}$ is in probability units. For sources see Table 7.3

Table 7.5 Impact of standard deviation change in county characteristics on probability of yes vote

County level variable	Change in probability
Value per voting age individual	
Logarithm of taxable wealth	−0.0036
% voting age pop doctors	−0.0055
% religious members Christian Scientists × 100	0.0005
% voting age pop union members	0.0040
% voting age pop illiterate	−0.0051
% voting age pop foreign-born	0.0028
% votes cast for Wilson in 1916	0.0073
% votes cast for Benson in 1916	0.0000
Value per voting age individual	
Crops	−0.0008
Manufacturing products	0.0009
Mineral products	−0.0008
Railroad assessments	−0.0010
Dummy=1 if county lumber producer	−0.0014
Dummy=1 if city of 25,000+	0.0011
No of hospital beds per capita	−0.0001

Note: Calculated from the regression results in Table 7.3. The probability of a yes vote is 0.17

Voters for Wilson and Benson were more likely to vote for the amendment than voters for Hughes. The percentage voting for Wilson was a positive predictor of the mean difference in utilities, suggesting that Progressive voters approved of the health insurance amendment. When a southern dummy is added to the regressions, its coefficient is significantly negative and reduces the explanatory value of the voting variables because the south did not vote for Wilson. The percentage voting for Benson was a significant, negative predictor of voter abstention and thus a significant, positive predictor of the perception threshold. The perception threshold was also significantly and positively related to literacy and negatively related to foreign birth.

The quantitative assessment of voting on the 1918 California health insurance referendum that has just been presented suggests that the persuasiveness of interest groups such as doctors and to a lesser extend trade unions did matter. Political affiliation was another important factor. But, none of the variable could fully account for the defeat of health insurance. The defeat was simply too resounding. A standard deviation change in any of the variables would not have affected the outcome (Table 7.5).

Although it was not possible to examine the impact of all interest groups, such as insurance companies, had these groups made doctors twice as effective, they still would not have affected the outcome. Even if voter turnout had been higher than 59%, a low turnout attributed by newspapers to war conditions, “the lack of pep on the part of candidates” (*Los Angeles Times*, October 13), and the influenza

epidemic which frightened voters away from the polls (*Oakland Tribune*, November 5), the outcome would not have changed.[28] Although the fraction of registered voters showing up at the polls was higher in counties with higher proportions of doctors, union members, and Christian Scientists, suggesting that interest groups may have been able to get the vote out, the impact of these variables on voter turnout was small and insignificant. Even if more individuals had registered to vote so that more than 42% of all native-born or naturalized foreign-born individuals age 21 or over had gone to the polls, this too would not have affected the outcome. None of the variables examined were significant predictors of turnout, suggesting that turnout bias is not a problem.

7.5 Explaining the Defeat

The resounding defeat of health insurance, a defeat so complete that it could not be explained by any of the included variables, suggests that explanations that have focused on the campaigns of interest groups have been too narrow. The previous investigation of private health insurance suggests that one plausible explanation for the defeat is simply that there was very little demand for health insurance. Health insurance was of limited use, in part because individuals could self-insure against the risk of lost wages either through the wages of other family members, through savings, or through charity. The lack of demand for private health insurance may not have been the only factor. Other possible explanations for the defeat of the amendment include the impact of WWI, popular disgust with regulatory politics, and anti-tax sentiment.[29]

The relative importance of these factors in the defeat of state provided health insurance can be assessed from an examination of newspaper editorials and advertisements shortly before the election. The newspapers that I examined were the *Sacramento Bee*, the *San Jose Mercury Herald*, the *Oakland Tribune*, the *San Francisco Bulletin*, the *San Francisco Chronicle*, the *San Francisco Examiner*, and the *Los Angeles Times*. The *Sacramento Bee* urged a yes vote for the measure, on the grounds that it was purely an enabling amendment. The *San Jose Mercury Herald*, the *San Francisco Bulletin*, and the *San Francisco Examiner* printed no editorial recommendations regarding health insurance.[30] The *San Francisco*

[28] In contrast, during the 1916 Presidential election, voter turnout was 80%. The number of registered voters was obtained from California Registrar of Voters (1918).

[29] Skocpol (1992) has argued that the failure of the Progressive agenda can be traced to dissatisfaction with the federal pension program run for the benefit of Union Army veterans, but I was able to find no mention of the Union Army pension program in any of these debates. Supporters of state-provided health insurance cited the program that served WWI soldiers and which grew out of the Union Army pension program Ohio Health and Old Age Insurance Commission (1919).

[30] The *San Francisco Examiner* considered prohibition the only issue of the campaign. It was in favor.

Examiner printed side by side pro and con articles. The *Oakland Tribune*, the *San Francisco Chronicle*, and the *Los Angeles Times* were opposed. Although all three newspapers noted the German origin of health insurance, only the *Los Angeles Times* emphasized this feature, urging the defeat of health insurance as a patriotic duty and entitling an editorial "Swat the Huns on the Ballot" (*Los Angeles Times*, October 27, 1918). In fact, anti-German sentiment was already strong when the California legislature passed the amendment and the treaty of Versailles had not yet been signed when a health insurance bill passed the New York State Senate. In contrast to the *Los Angeles Times*' campaign against health insurance, the *San Francisco Chronicle* entitled its voting recommendations "Guide Will Assist Voters to Defeat Tax Eater Hordes" (*San Francisco Chronicle*, November 5, 1918), emphasizing the cost of the program, that benefits would be given to a small class of people, and that a new commission with unlimited powers would be created. The *Oakland Tribune* primarily emphasized the cost of the program.

Two advertisements in favor of health insurance were run. Both of them mentioned Senator Hiram Johnson's endorsement of the measure and both of them appealed to workers' familiarity with workmen's compensation by stating that health insurance "does for sickness what workmen's compensation does for industrial accidents. It includes the family." Ads opposing health insurance generally took the form of a complete list of ballot recommendations. These were paid for by the San Francisco Chamber of Commerce, the Civic League, the Legislative Committee of the San Francisco Real Estate Board, and the Governing Committee of the Los Angeles Realty Board. Only the Civic League recommendations gave any reason for voting against health insurance and that was "Creation of new commission with unlimited powers. Makes insurance compulsory; creates an unlimited tax."

The *San Jose Mercury Herald* ran advertisements opposing health insurance that specifically singled out the amendment. Ads appeared listing San Jose dentists, doctors, attorneys, and bankers who opposed the health amendment. An ad run November 5 and 6 pointed out the quality of care would suffer because "The care and attention of insurance physicians will be perfunctory and will lack that personal interest which is so necessary to the patient." Doctors had already had unfavorable experiences with insurance under workers' compensation laws. The commercial insurance companies that provided accident insurance often contracted with physicians in advance to care for the injured and these contracts may have encouraged doctors to provide less than adequate care (Numbers 1978, p. 114). The advertisement in the *San Jose Mercury Herald* went to add,

> "This Prussian device will raise California taxes $15,000,000 in one year."
> "It will require $50,000,000 a year to run the Health Insurance business – you will pay it all."

On November 5 there was also an ad in the form of a long article entitled "Compulsory Insurance a Failure in England. No Improvement in Public Health. Medical and Pharmaceutical Professions Are Affected." The advertisement campaign is summarized in Table 7.6.

There were no advertisements against health insurance in either the *Oakland Tribune* or the *Sacramento Bee*. In the *San Francisco Bulletin*, the *San Francisco*

Table 7.6 Number of advertisements relating to health insurance before 1918 election

Newspaper	Favoring	Opposing		
		Total	Ballot list	Individual
Los Angeles Times	1	1	1	0
Oakland Tribune	2	0	0	0
Sacramento Bee	1	0	0	0
San Francisco Bulletin	2	4	4	0
San Francisco Chronicle	2	3	3	0
San Francisco Examiner	2	8	8	0
San Jose Mercury Herald	1	4	0	4

Note: The advertisements favoring health insurance appeared in the *Los Angeles Times* on November 4, in the *Oakland Tribune* November 2 and 3, in the *Sacramento Bee* on November 2, in the *San Francisco Bulletin* on November 2 and 4, in the *San Francisco Examiner* and *San Francisco Chronicle* November 3 and 4, and in the *San Jose Mercury Herald* on November 5. The advertisements that were complete ballot recommendations appeared in the *Times* on November 4, in the *Bulletin* on November 2 and 4, and in the *Chronicle* and *Examiner* on November 2, 3, and 4. The advertisements in the *Mercury Herald* were run October 24 and November 4 and 5

Chronicle, the *San Francisco Examiner*, and the *Los Angeles Times*, the only advertisements against the health insurance took the form of complete lists of ballot recommendations. Only the *San Jose Mercury Herald's* opposing ads were specifically targeted towards the health insurance amendment. Thus, just before the election, the pro health insurance forces do not appear to have been outspent in the advertising campaign.[31]

The other information that voters received about health insurance just before the general election was in newspaper reports of a public hearing on health insurance organized by the California Social Insurance Committee on October 22 in San Francisco. The *San Francisco Examiner*, the *San Francisco Chronicle*, and the *San Francisco Bulletin* reported on the public meeting, emphasizing the alien nature of health insurance, its cost, and its failure in Britain, respectively. On November 5 the *San Jose Mercury News* and the *Los Angeles Times* ran articles outlining the stand of Frederick Hoffman of Prudential Insurance which dwelt largely on the failure of state provided health insurance to improve public health.

What the editorials, advertisements, and articles suggest is that although there was anti-German hysteria, disgust with commission politics, and uncertainty over exactly what type of bill would result from a yes vote for the amendment, cost was the primary concern. A huge bureaucracy would need to be created and the cost to the State would be "nearly the entire present cost of the strictly government expenses of the State" (*San Francisco Chronicle*, October 25). The need for such a large expenditure was unclear. The weak relationship between income and

[31]The outspending hypothesis has been advocated by Viseltear (1969). However, the anti-health insurance forces did have the benefit of newspaper editorials and may have already felt that they had public opinion on their side.

health insurance and income and medical care suggests that access to either health insurance or medical care does not appear to have been a problem. Both health insurance and medical care were of rather limited use. There was much evidence that state provided health insurance would not improve public health and that only those who currently had health insurance, a relatively small group, would actually benefit. Those who preferred self-insurance would be faced with a large expense.

7.6 Concluding Remarks

This paper has analyzed the demand for both private and state-provided health insurance. In the case of private health insurance, I showed that both health insurance and medical care were of limited use and that the relationship between income and health insurance and income and medical care was relatively weak, suggesting that money could buy little in the way of improvements in medical care. These results implied that there should be very little demand for state-provided health insurance and indeed there was not. Although the persuasiveness of interest groups such as doctors and to a lesser extend trade unions did contribute to the defeat of state-provided health insurance matter, none of the variables could explain such a resounding defeat. Evidence from newspaper editorials, advertisements, and articles suggested that the absence of consumer demand for health insurance together with concerns over the cost of state-provided health insurance defeated the measure. My findings are in contrast to those of other researchers who have emphasized the role of a politically powerful medical profession and of World War I.

The role of cost considerations in the defeat of the health insurance measure suggests that even before World War II voters may have behaved as "fiscal conservatives" (Peltzman 1992). The greater centralization of power in Europe and the absence of referenda may explain why state-provided health insurance was passed in these countries.[32] Using recent data, Matsusaka (1995) finds that state spending is higher in states that do not use initiatives and concludes that the initiative leads to a reduction in the overall size of government and limits the level of distributional activities. Fishback and Kantor (1994) argue that complex laws like workers' compensation often faced opposition in referenda because of the magnitude of the proposed changes and the dearth of information with which voters could form expectations. They find that the more extreme components of workers' compensation laws, such as generous benefits, state insurance, or an expanded bureaucracy, had to be weakened or eliminated in order to win voters' support. State-provided health insurance would have resulted in a far vaster bureaucracy and have

[32] Alternatively, Lindert (1994) argues that the United States lagged behind British social spending because the United States was younger, had a lower voter turnout, and was a country in which middle class incomes and sympathies were closer to the top.

cost much more than any proposed workers' compensation program and would have been even more of a "plunge in the dark" for voters.[33]

Why weren't state provided health insurance plans modified until a plan could win voter support? After World War II this is indeed what happened with the provision of health insurance to the poor and the old. But, the primary aim of private health insurance in 1918 was to provide cash benefits to employed, prime-aged males. All that state-provided health insurance would have done would have been to change the financing of the system, not its actual structure. Coverage would have been limited, as noted in editorials opposing health insurance. Had benefits been restricted, coverage through the private sector would have been even more attractive than the proposed coverage through the government sector and voter support for state-provided health insurance perhaps even smaller.

Acknowledgements I have benefited from the comments of Herbert Emery, Claudia Goldin, Matthew Kahn, Robert Margo, Peter Temin, and participants at the 1995 Cliometrics Conference and the 1995 NBER Health Care Summer Institute. I gratefully acknowledge the support of a NIA Aging Fellowship at the National Bureau of Economic Research in the 1995–1996 academic year.

References

American Medical Association (1918) American medical director, 1918. American Medical Association, Chicago

Anderson OW (1951) Compulsory medical care insurance, 1910–1950. Ann Am Acad Pol Soc Sci 273(1):106–113

California Bureau of Labor Statistics (1892) A survey of 3,493 wage earners in California, 1892. In: Fifth biennial report of the Bureau of Labor Statistics of the State of California. Historical Labor Statistics Project, Institute of Business and Economic Research, UC-Berkeley. California State Printing Office, Sacramento

California Bureau of Labor Statistics (1918) Eighteenth biennial report of the Bureau of Labor Statistics of the State of California, 1917–18. California State Printing Office, Sacramento

California Registrar of Voters (1918) Statement of vote. California State Printing Office, Sacramento

California Secretary of State (1913) California blue book or state roster, 1911. California State Printing Office, Sacramento

California Social Insurance Commission (1917) Report of the Social Insurance Commission of the State of California. California State Printing Office, Sacramento

California State Board of Equalization (1918) Report of the California State Board of Equalization for 1917–1918. California State Printing Office, Sacramento

California State Tax Commission (1917) Report of the State Tax Commission of the State of California. California State Printing Office, Sacramento

Clubb JM, Flanigan WH, Zingale NH (2006) Electoral data for counties in the United States: presidential and congressional races, 1840–1972. Inter-university Consortium for Political and Social Research, Ann Arbor

[33] A "plunge in the dark" is how Florence Kelly, General Secretary of the Consumers' League and official representative of women wage earners, described compulsory insurance (cited in Hoffman (1917)).

Deacon R, Shapiro P (1975) Private preference for collective goods revealed through voting on referenda. Am Econ Rev 65(5):943–955
Dewhurst JF, Associates (1955) America's needs and resources. Twentieth Century Fund, New York
Emery JCH (1993) The rise and fall of fraternal methods of social insurance: a case study of the independent order of oddfellows of British Columbia sickness insurance, 1874–1951. PhD thesis, University of British Columbia
Fishback P, Kantor S (1994) Coalition formation and the adoption of workers' compensation: the case of Missouri, 1911 to 1926. In: Goldin C, Libecap G (eds) The regulated economy. University of Chicago Press, Chicago
Hoffman FL (1917) Facts and fallacies of compulsory health insurance. Prudential Press, Newark
Illinois Social Insurance Commission (1919) Report of the health insurance commission of the state of Illinois. Illinois State Journal Company State Printers, Springfield
Inter-University Consortium for Political and Social Research (2005) Historical, demographic, economic, and social data: the United States, 1790–1970 (ICPSR 3). Inter-university Consortium for Political and Social Research, Ann Arbor
Lindert PH (1994) The rise of social spending, 1880–1930. Explor Econ Hist 31(1):1–37
Matsusaka JG (1995) Fiscal effects of the voter initiative: evidence from the last 30 years. J Polit Econ 103(3):587–623
Naylor D (1986) Private practice, public payment: Canadian medicine and the politics of health insurance, 1911–1966. McGill-Queen's Press, Kingston
Numbers RL (1978) Almost persuaded: American physicians and compulsory health insurance, 1912–1920. Johns Hopkins University Press, Baltimore
Ohio Health and Old Age Insurance Commission (1919) Health, health insurance, old age pensions. FJ Heer Printing, Columbus
Peltzman S (1984) Constituent interest and congressional voting. J Law Econ 27(1):181–210
Peltzman S (1992) Voters as fiscal conservatives. Q J Econ 107(2):327–361
Poole KT, Rosenthal H (1993) The enduring nineteenth-century battle for economic regulation: the Interstate Commerce Act revisited. J Law Econ 36(2):837–860
Rogin MP, Shover JL (1970) Political change in California: critical elections and social movements, 1890–1966. Greenwood Publishing Corporation, Westport
Rosen G (1977) Contract or lodge practice and its influence on medical attitudes to health insurance. Am J Public Health 67(4):374–378
Rubinow IM (1916) Standards of health insurance. H Holt and Company, New York
Shonick W (1995) Government and health services: government's role in the development of US health services, 1930–1980. Oxford University Press, Oxford
Skocpol T (1992) Protecting soldiers and mothers. Harvard University Press, Cambridge
Starr P (1982) The social transformation of American medicine: the rise of a sovereign profession and the making of a vast industry. Basic Books, New York
United States Bureau of the Census (1906) Census of religious bodies. Government Printing Office, Washington
United States Bureau of the Census (1926) Census of religious bodies. Government Printing Office, Washington
United States Department of Labor, Bureau of Labor Statistics (1985) Cost of living in the United States, 1917–1918 (ICPSR 8399). Inter-university Consortium for Political and Social Research, Ann Arbor
Viseltear AJ (1969) Compulsory health insurance in California, 1915–1918. J Hist Med Allied Sci 24(2):151–182
Warren BS, Sydenstricker E (1916) Health insurance: its relation to public health. United States Public Health Service, Washington
Whaples R, Buffum D (1991) Fraternalism, paternalism, the family, and the market: insurance a century ago. Soc Sci Hist 15(1):97–122

Chapter 8
What Determines the Allocation of National Government Grants to the States?

John Joseph Wallis

Abstract During the New Deal the federal government initiated a policy of massive grants to states for support of social welfare and other programs. Since that time grants have come to be an integral part of the American fiscal system, and scholars have continued to debate whether the allocation of federal grants between the states is motivated primarily by political or social and economic objectives. This paper shows that, during the 1930s, both political and economic effects were important determinants of grant allocation, but that the Congressional factors considered by Anderson and Tollison are not important while the Presidential factors considered by Wright are. When the analysis is extended to the years 1932–1982, however, Congressional influences do seem important. On the other hand, the dominant influence on federal grant policy over the larger sample appears to be state government expenditures, while both political and economic influences play a smaller role.

8.1 Introduction

One of the most prominent changes in the structure of the American fiscal system in the twentieth century is the growing importance of federal grants-in-aid to state and local governments. At the turn of the century federal grants to state and local governments were $7 million, too small to warrant separate treatment in the Annual Report of the Secretary of the Treasury. In 1992, federal grants totaled $171,400 million. They account for 11.1% of total government expenditures (17.5% of non-military expenditures) and 20% of state and local revenues.[1] As important as their fiscal magnitude, intergovernmental grants have become an integral part of

[1]Figures taken Advisory Council of Intergovernmental Relations (1994).

J. J. Wallis (✉)
University of Maryland, College Park, MD, USA
e-mail: wallis@econ.umd.edu

J. Hall, M. Witcher (eds.), *Public Choice Analyses of American Economic History*,
Studies in Public Choice 35, https://doi.org/10.1007/978-3-319-77592-0_8

financing expenditures for education, welfare, and transportation infrastructure. The current debate over the size and role of the federal government necessarily involves rethinking the fiscal structure of all three levels of government, since major changes in federal government expenditures are only possible through major changes in grant policy. Welfare reform, for example, is as much about what level of government will tax and spend as it is about how much will be spent.

While the economic theory of intergovernmental grants is well developed, the empirical study of the allocation of grants has focused as much on politics as on economics.[2] Political factors are almost always important determinants of the allocation of federal grants between states, and often are quantitatively more important determinants of grant policy than economic variables like income or urbanization. Whether a state is critical to a President's re-election or is represented by a powerful Congressman appears to affect grants as much or more than whether a state has low income, and therefore should qualify for larger redistributive grants, or has a larger urban population, and therefore should qualify for larger urban targeted grants. In fact, states with higher incomes often appear to receive larger grants. These conclusions suggest that the enormous growth in federal government grants may not be the results of attempts at fiscal rationality, but may instead be an enormous port-barrel boondoggle imposed on the American taxpaying public for the political benefit of incumbent Presidents and Congressmen.

The results have a direct impact on historical and current debates. The allocation of federal government grants during the New Deal administration of Franklin Roosevelt has received a great deal of attention.[3] The focus on the New Deal is understandable, since expansion of federal grants were central to New Deal programs in agriculture, social welfare, and infrastructure investment. Federal expenditures grew from $4.3 billion in $10.1 billion in 1940, while expenditures in "cooperatively administered programs" rose from $0.2 billion to $3.9 billion, two thirds of the increase.[4] Our view of the New Deal is much different if these grant programs, which represent the cutting edge of liberal social reform in American history, were driven primarily by partisan political concerns.[5]

The current debate over the size and structure of the federal government is intimately tied up with the nature and operation of grants. Federal expenditures for welfare, education, and transportation are made primarily through intergovernmental grants. Since the Reagan administration there has been an ongoing debate about

[2]The public finance literature is reviewed in Oates (1972) and Gramlich (1977). For specific studies see Bungey et al. (1991), Chernick (1979), Yinger and Ladd (1989), Craig and Inman (1982), Ladd (1991), and Holcombe and Zardkoohi (1981).

[3]The New Deal literature includes Arrington (1969, 1970), Reading (1973), Wright (1974), Wallis (1984), Wallis (1987), and Anderson and Tollison (1991). The structure of New Deal programs is considered in Wallis (1991).

[4]See Wallis (1984) for a closer analysis of the grant numbers and the problems with using the census numbers.

[5]There is a larger historical literature on grants and intergovernmental relationships during the New Deal (Clark 1938; Benson 1941; Bittermann 1938).

whether these functions should be returned to the states. Arguments for keeping some measure of administrative control and financial discretion at the national level are considerably weaker if it appears that national control serves ends that primarily political. This is even more true if grants that are, in a general way, designed to redistribute income and help urban areas, are less likely to go to states with low incomes and large urban populations.

Advances in the study of grant allocation have focused on the political side. Anderson and Tollison explored the use of a new set of variables that represent Congressional politics during the New Deal. This paper also expands the set of political variables that affect grants. The existing set of political variables are designed to pick up "swing" states. One might expect that a Democratic President would direct large amounts of funds to states that traditionally vote Democratic. But shifting the vote 1% toward the Democrats in a state where Democrats traditionally receive 70% of the vote is much less productive than an equivalent grant and 1% vote shift would be in a state where Democrats traditionally receive 49% of the vote. While the logic is sound, over the long term it implies that the Democrats (or Republicans) systematically ignore their strongest supporters in favor of states that more often support their opponents. This logic underlies the construction of Wright's "political productivity" variables as well as Anderson and Tollison's focus on key Congressional leaders who can swing the legislative process. Using data covering years from 1932 to 1982, it is impossible to distinguish between the response of grant allocation to the level of political support in a state as well as the importance that each state possesses as a potential swing state.

The major emphasis of the paper, however, is not on improving the set of useful political variables. The econometric specifications used in most studies look at simple cross sections of states and focus solely on the behavior of the federal government. Many grant programs, however, contain explicit or implicit matching provisions to provide state governments with an incentive to supplement grant moneys with their own tax dollars. Consideration of the simultaneity of state spending decisions and federal grant allocations is very important. Federal grant allocations also have significant idiosyncratic state-specific component. When state fixed effects are controlled for, the empirical results change dramatically. This paper uses a more appropriate econometric specification to ascertain what determines the allocation of federal grants. I find that both political and economic factors matter statistically; that the economic factors exert the right kind of influence on grants; and that quantitatively the effects of both political and economic factors are rather small. By contrast, the effect of state expenditures on federal grants is statistically significant and quantitatively large. Variations in state expenditures explain a large part of the variance in federal grants across states.

After considering a few methodological issues, I briefly re-examine the conclusions of the New Deal literature using a comparable data set to re-estimate Wright (1974), Anderson and Tollison (1991), and my earlier work (Wallis 1991). A full investigation is then done on a long panel data set, including information from the 1932, 1942, 1962, 1972, and 1982 *Census of Governments*. Unlike the 1930s data, all of the economic and political variables in the large data set vary over states and

time, and complete state and local fiscal information is available for each year. With the twentieth century data set we can draw firm conclusions about the relative effect of economic, political, and state fiscal variables on the allocation of national grants between states.

8.2 Some Methodological Considerations

Our model of the grant process pictures federal governments structuring grant allocations to maximize the value of those grants to themselves and/or their constituents. Congress decides the formulas under which some grants will be allocated and the executive branch has discretionary control in the awarding of other grants. This translates into an estimating equation of the form:

$$FG_i = \alpha + \beta E_i + \gamma P_i + \varepsilon \tag{8.1}$$

where FG_i is per capital federal grants to state I; E and P are vectors of economic and political variables; and the error term, ϵ_t, is a normal white noise error.

Equation (8.1) is not, however, an accurate representation of how the grant allocation process actually works. Grants are typically allocated through a series of formulas that include factors like population, income, or highway mileage, as well as measures of state (or local) financial participation.[6] Congress may have in mind an allocation of grants that reflect economic and political factors, but state and local governments have independent input into the allocation process. At one extreme, with strict matching, federal grants are simply a linear function of state expenditures. This was true for all of the categorical assistance programs put in place in the Social Security Act of 1935, and is still true today. Many grants fall short of strict, open-ended matching formulas, but many involve some attempt to reward states that make a larger effort to cooperate with the federal goals by rewarding those states with larger grants.[7]

States also differ in ways not capture by economic or political variables, or captured only imperfectly. These persistent differences between states produce differences in state fiscal activity, as well as persistent differences in grants. These "fixed effects" can be controlled for if we have a panel data set that varies over time.

[6]Both state and local governments are important in this process. But state government's received the largest share of federal grants and, as the empirical sections will demonstrate, it is state rather than local expenditures that influence the allocation of federal grants.

[7]There is little alternative for the Congress and the federal grant administrators. With the exception of general revenue sharing, grant programs are intended to increase government provision of particular services. If states are allowed to reduce their expenditures by substituting federal grants for state dollars, then grants will have a smaller effect of total expenditures than if the grant structure creates incentives for larger state expenditures. See Chernick (1979) for an interesting examination of the interaction between federal and local officials.

A more appropriate set of estimating equations are:

$$FG_{it} = \alpha_g + \beta E_{it} + \gamma P_{it} + SX_{it} + \epsilon_i + \epsilon_t + \epsilon \tag{8.2}$$

$$SX_{it} = \alpha_x + \lambda Z_{it} + \psi NG_{it} + SX_{it} + \epsilon_{si} + \epsilon_{st} + \epsilon_s \tag{8.3}$$

where subscripts i refer to states and t refers to time, SX is state expenditures, Z is a vector of variables explaining state expenditures, and $\epsilon_s' s$ are the errors for the state expenditure equation.[8]

If Eq. (8.1) is estimated when Eqs. (8.2) and (8.3) are true then Eq. (8.1) estimates will suffer from several defects. One will be omitted variable bias. This may be a problem for all variables, but it turns out to be particularly important for some of the political variables. The political variables tend to be stable over significant periods of time and are correlated with ϵ_i and ϵ_t. The coefficients on the political variables, therefore are sensitive to the specification of the error term.

Second, even if state and local expenditures are included as economic variables in Eq. (8.1), the estimates will suffer from simultaneity bias. This turns out to be important for a number of coefficients, the most important being per capita income. The professed goal of many federal grant programs is to redistribute income. But in the United States, the bulk of social welfare spending is under the direct control of the states, financed by matching grants from the federal government. If states with higher incomes tend to spend more on social welfare, then estimates of Eq. (8.1) will tend to show that states with higher incomes receive larger grants, in direct contradiction to the stated goals. One of the more persuasive pieces of evidence supporting the case that political factors are more important than economic factors is that some economic variables, like income, appear to have a perverse effect on grants. If this result is, however, an artifact of the estimation technique, then support for the political factors argument is considerably reduced.

8.3 The New Deal

Leonard Arrington was the first to notice a peculiarity in the allocation of New Deal expenditures across states.[9] They tended to be highest in states with relatively high incomes in the West, and lower in states with relatively low incomes in the

[8] Local government expenditures can be treated separately from state expenditures, as will be done in the empirical section.

[9] New Deal "expenditures" are the sum of grants made to states and expenditures within states made directly by the federal government. For many New Deal programs the distinction between direct expenditures and grants was not a clear one. For example, under FERA, relief grants were made to the states. The states then distributed the grants to local relief agencies who distributed the funds to individuals. Under the WPA, which was FERA's primary successor in 1935, state and local WPA agencies and employees were actually federal government agencies and employees. So WPA wages were given "directly" to individuals while FERA relief benefits were granted to

South, despite the fact that the South was solidly Democratic. If Roosevelt were rewarding his supporters the Southerners should have received the largest grants per capita, not the smallest grants (Arrington 1969, 1970). Reading implemented an econometric analysis of New Deal grants using a report published by the Office of Government reports, Statistical Section in 1940 (Office of Government Reports, Statistical Section 1940). He found support for the idea that New Deal expenditures were intended to promote the natural resources of the country (as measured by the percentage of land in a state owned by the federal government) and to compensate for the effects of the Depression, but found no support for the idea that New Deal programs promoted reform. Wright (1974), using Reading's data, tested the hypothesis that political considerations were an important part of the allocation process. Wright found that his political variables did a much better job of explaining the patterns of expenditures across states than did the economic variables.

Because of data limitation, both Reading and Wright focused on the federal government and limited themselves to one cross-section of 48 states, covering the entire period from 1933 to 1939. Most of the New Deal programs were intergovernmental, not national, and many involved explicit or implicit matching grants. Unfortunately, state and local government expenditures are not available over the years 1933–1936. I showed that federal grant policy was responsive to state and local spending after 1936 (Wallis 1984). In a second paper, I incorporated information on annual state-level employment and annual per capita income to create a panel of seven years and forty-eight states (Wallis 1987).[10] Using the panel data set, it was apparent that both economic and political factors had an effect on New Deal policy. On balance it appeared that economic effects may have been more important quantitatively than the political effects, but the total effect on grants was relatively small. Neither political or economic factors alone, as proxied by employment and Wright's political productivity measures, drove New Deal allocation policy.

Anderson and Tollison (1991) have suggested that Wright and I erred in our characterization of the political process. Wright's political measures are built around voting returns from presidential elections. Congress, however, exercises an important role in the allocation of federal grants and spending through the choice of allocation formulas and control over agency funding. They incorporated several variables that proxied for the political influence of Congressional members—key leadership positions, tenure in office, and tenure on the appropriations committees—and found that these Congressional political variables had an important effect on the allocation of New Deal expenditures.

All of these papers use Reading's data set, taken from the Office of Government Reports documents. But inconsistencies in how variables are measured or calculated

the states. As a result, both the terms expenditures and grants are used to describe the New Deal allocations. FERA policies are described in Wiliams (1939) and WPA policies in Howard (1943).

[10]The employment index is presented and explained in Wallis (1989).

make direct comparisons of their results difficult. The replicated estimates that follow utilize a common data set, based on Reading's original data and the following variable definitions.

The economic variables include the proportion of the population living on farms; the change in per capita income, 1929 to 1933 (measured as a % of 1929 income; the percentage of land owned by the federal government in each state; unemployment rates in 1937; and total farm value per capita in 1929 including land, buildings, implements and machinery, and livestock.

The political variables are broken into two groups, the presidential variables and the congressional variables.

The presidential variables were constructed by Wright. He collected information on the Democratic share of the presidential vote from 1896 to 1932, and fitted a linear trend to the data to "predict" the vote in 1936. Then using information on the standard deviation around the trend, he constructed a voting distribution for each state, with the predicted mean and the observed standard deviation. He then evaluated the probability that the Democratic vote would be 50% or greater, Pr_0, and compared that to the probability that the vote would be 50% or greater if the whole distribution was shifted 1% towards the Democrats, Pr_1. The change in the probability of a Democratic victory was then multiplied by the number of electoral votes a state controlled, and divided by 1% of the voting population.[11] Wright also included the standard deviation of the Democratic share of the presidential vote over the preceding ten elections as a measure of how easy it might be to influence voters in a state, arguing that voters in states with higher standard deviations might be easier to influence through grants.

I constructed an alternative version of Wright's political productivity index for each state in 1932, 1942, 1962, 1972, and 1982. I took the Democratic share of the vote in the ten preceding presidential elections, and calculated the mean and standard deviation of that sample. Assuming a normal distribution, I then calculated the difference in the probability of receiving 49 and 50% of the vote, $Pr_{0.01} = Pr_{0.49} - Pr_{0.50}$. Then I multiplied by the number of electoral votes and divided by population to get the political productivity measure.

Anderson and Tollison do not use Wright's measures. The heart of Anderson and Tollison's argument is that key Congressional leaders should have been able to obtain larger grants for their constituents. They used six measures, three each for the Senate and the House. In the Senate these are: a dummy variable if a state's Senator was President Pro Tem or Senate Majority Leader during the period from 1933 to 1939, the length of consecutive tenure in months of the state's Senate delegation in December 1937, and the length of consecutive tenure in month of Senators serving on the Senate Appropriations Committee in December 1937. The House variables are the same, with the exception that the House leadership positions are Speaker of the House and Majority Leader. The also include electoral votes per capita, the

[11] See Wright (1974) for a complete description.

state's rank in electoral votes per capita, and Roosevelt's share of the total vote in the 1932 election. I have made some changes to their data to fix some apparent errors and to facilitate comparison with the rest of the century.[12]

There is, however, a serious problem with Nevada. The Senate leadership variable is based on whether a state's Senator was the majority leader or President Pro Tem. Since Key Pittman of Nevada was President Pro Tem for the entire period, Nevada gets a 1 in the dummy variable for Senate leadership. Unfortunately, Nevada also receives far and away the largest per capita grants. In Reading's total allocations Nevada receives $1499.39, fully one half again as much as the next state, Montana with $986.30, and more than triple the national average. Not surprisingly, Senate leadership is consistently positive, both in Anderson and Tollison's results and in the panel results for the 1930s. Anything that proxies for Nevada will be positive. Anderson and Tollison also estimate several specifications including electoral votes per capita, which also picks up the Nevada effect. Electoral votes per 1000 population in Nevada is 0.039. The state with the next highest value is Delaware with 0.012. The state with the lowest value is New York with 0.0037. The national average is 0.006, with a standard deviation of 0.0044. Nevada is seven standard deviations from the mean, with three times the value of the next state, and five times the mean value.[13]

Table 8.1 reports variable means and standard deviations for the simple cross-section of 48 states used by Wright and by Anderson and Tollison, and the panel data with 6 years of data used by Wallis. The dependent variable in the regressions that follow is per capita expenditures as reported by Reading (1973), measured in 1935 dollars. Table 8.2 replicates the original specifications of Anderson and Tollison,

[12] Although Anderson and Tollison (1991) report that both Senate and House tenures are counted in months, their table of summary statistics on p. 170 gives the average tenure in the Senate as 187 months, while the average tenure in the House is 0.22 months, roughly 7 days. Clearly there was an error in their coding. Since the House tenure variables enter into every regression, my results are different than theirs, sometimes significantly so. I have dated tenure in months from the opening day of the Congressional session beginning January 3, 1937 rather than in December of 1937 to facilitate collection of the panel data for the 1930s as well as for the twentieth century. I find average Senate tenure of 176 months (compared to 187 in Anderson and Tollison (1991)), average House tenure of 582 months, average Senate Appropriations Committee tenure of 42 months (compared to 55 months), and average House Appropriations Committee tenure of 59 months.

Anderson and Tollison (1991) report a mean for their Senate Leadership variable of 0.04, implying that Senators from two states held leadership positions between 1933 and 1939 (2/48=0.04). In fact, three Senators held leadership positions. The President Pro Tem of the Senate was Key Pittman of Nevada throughout the period. Joseph Robinson of Arkansas was elected Majority Leader the 73rd, 74th, and 75th Congresses, spanning the years 1933 to 1939. But Robinson died on July 14, 1937 and was replaced by Alben Barkley of Kentucky on July 22. As a result, three states are included in the Senate Leadership dummy variable: Nevada, Arkansas, and Kentucky.

[13] The Nevada effect can be seen in Anderson and Tollison's (1991) original estimates. When they use electoral votes per capita, the Senate leadership variable is small and statistically insignificant. When they use the state's rank in electoral votes per capita, the Senate leadership variable is large and statistically significant.

Table 8.1 Variable means and standard deviations

Aggregate variables, 1933–1939 (N=48)	Mean	St. dev.
Per capita national spending	293.44	178.11
Percentage decline in income	−28.73	7.89
Unemployment, 1930	5.75	2.25
Unemployment, 1937	4.22	0.89
Farm value per capita	0.64	0.55
Farm population share	0.29	0.16
Percent federal land	13.45	20.63
Democratic vote, 1932	64.75	12.74
Per capita electoral votes	0.0060	0.0045
Rank of electoral votes	24.50	14.00
Senate leadership	0.063	0.24
House leadership	0.083	0.28
Senate tenure	175.90	121.97
House tenure	582.46	721.74
Senate appropriations	41.79	65.75
House appropriations	59.15	110.87
Political productivity	0.041	0.036
Standard deviation of vote	10.18	4.33
Annual variables, 1934–1939 (N=288)	Mean	St. dev
Per capita national spending	43.78	27.34
Percentage Decline in income	−0.29	0.08
Farm population share	0.29	0.16
Percent federal land	13.45	20.45
Senate leadership	0.042	0.20
House leadership	0.042	0.20
Senate tenure	559.60	697.96
House tenure	559.60	697.96
Senate appropriations	42.47	65.08
Political productivity	0.041	0.036
Standard deviation of vote	10.18	4.29
Employment index	85.98	9.84
Per capita real income	456.85	164.99
Lagged per capita spending	37.77	28.36

Wright, and Wallis specifications in columns (1), (2), and (3). In order to facilitate comparisons, all replications use the same data definitions. Differences between the original and replicated results are due to different scales for some variables, different measures of income decline between 1929 and 1933, and minor improvements made to the employment index and to the spending variables since the original studies were published. Columns (4) and (5) of the table present a full panel specification. Grants and the employment index are treated as endogenous variables to account for the possibility that more relief funds, and therefore more grants, would go to areas

where employment was lower.[14] These regressions include an annual state-level employment index, annual real per capita income, lagged real per capita spending, and drop the unemployment variable. To account for the Nevada effect, column (5) of the table excludes Nevada.

Table 8.2 Regression estimates

	A&T rep.	Wright rep.	Wallis rep.	Panel	Panel W/O nev.
Variable	(1)	(2)	(3)	(4)	(5)
Political productivity		1277.34	92.19	58.41	33.99
		(2.63)	(3.48)	(2.03)	(1.49)
Standard deviation of vote		11.15	0.93	0.55	0.52
		(2.28)	(3.47)	(1.89)	(2.31)
Senate leader	169.82			9.25	-2.05
	(2.98)			(2.02)	(0.48)
House leader	−29.84			−2.16	−3.21
	(0.56)			(0.55)	(1.04)
Senate tenure	−0.21			−0.01	−0.01
	(1.44)			(1.87)	(1.86)
House tenure	0.0279			−0.0004	−0.0009
	(0.95)			(0.34)	(0.95)
Senate appropriations	0.26			0.010	−0.013
	(1.10)			(0.76)	(1.27)
Income decline	0.27	−0.20	10.80	5.00	−9.24
	(0.11)	(0.07)	(0.70)	(0.32)	(0.72)
Unemployment	10.62	−6.05			
	(1.09)	(0.33)			
Farm value	127.57			0.56	0.32
	(3.39)			(3.11)	(2.21)
Farm population		118.71	63.18	44.60	38.06
		(0.85)	(3.68)	(2.61)	(2.81)
Federal land	4.96	4.97	0.14	0.22	0.18
	(5.12)	(4.87)	(1.79)	(2.76)	(2.85)
Employment index			−0.79	−0.49	−0.38
			(3.46)	(2.24)	(2.17)
Real income			0.043	0.018	0.019
			(3.08)	(1.31)	(1.76)
Lagged spending			0.55	0.53	0.56
			(13.27)	(13.04)	(15.08)

(continued)

[14]The instrument for the employment index in each state and year is a composite index of state employment constructed from occupational weights from the 1930–1940 census, and monthly information on national employment by industry from the Bureau of Labor Statistics and the Federal Reserve Board. See Wallis (1987) for a detailed explanation.

Table 8.2 (continued)

	A&T rep.	Wright rep.	Wallis rep.	Panel	Panel W/O nev.
Variable	(1)	(2)	(3)	(4)	(5)
Electoral votes					
Rank	3.20 (1.84)				
FDR vote	2.57 (1.61)				
Relief cases		0.00004 (0.00)			
Constant	−151.62 (0.82)	45.22 (0.42)	40.98 (2.48)	31.37 (1.97)	21.29 (1.70)
Observations	48	48	288	288	288
R-Squared	0.81	0.73			
F-Stat	14.2	15.9	129.1	84.6	77.1
Method	OLS	OLS	2SLS	2SLS	2SLS

Note: Dependent variable in each regression is real per capita state grants. Absolute value of t-statistics in parentheses

A direct test of Anderson and Tollison's hypothesis that congressional political factors influence grants against Wright's hypothesis that presidential political factors influence grants is reported in columns (4) and (5). Nevada is critical. Excluding Nevada from the data set eliminates the coefficient and significance of Senate leadership, the only Congressional political variable with any statistical power. Excluding Nevada also weakens the effect of Wright's political productivity variable, but the standard deviation in Presidential voting still remains positive and significant. It is difficult to find strong support for the Anderson and Tollison hypothesis in these results.

The regressions enable a direct test of the New Dealer's claim that funds went to states with lower employment (higher unemployment). Neither Wright nor Anderson and Tollison found that higher unemployment led to larger grants (columns (1) and (2)). But as I found earlier, when a panel data set is used, higher employment does lead to significantly smaller grants. States with higher incomes still receive higher grants, contradicting any suggestion that the New Deal was giving more money to poorer states. But state expenditures rise when state incomes rise, and therefore matching grants also rise, and we are unable to control for that simultaneity in this data set.

In general, the regressions suggest that economic factors were more important determinants of grants than political factors. Grants rise significantly with farm population, farm value, and income, and fall significantly as employment rises. A one standard deviation improvement in employment would reduce grants by about 8.5% ($3.74 per capita per year), while a one standard deviation increase in the political productivity index would raise grants by about 3% ($1.22) and a one standard deviation increase in the variability of voting would raise grants by about 5% ($2.23). When a state's Senator obtains a leadership post grants increase by 21%

($9.25 per year per capita), but only if you happen to live in Nevada. None of the other Congressional influence variables have much impact on spending , and the only statistically significant result is that states with longer tenured Senators receive slightly smaller grants.

In short, Anderson and Tollison's conclusion that congressional political interests were important rests firmly on Key Pittman and the anomalous experience of Nevada. Wright's conclusions that political factors were more important determinants of grants than economic factors is unwarranted, although the presidential political factors he identified still have explanatory power. If we want to disentangle these issues further we need better data.

8.4 Beyond the New Deal

While the New Deal was a fundamental turning point in the history of intergovernmental grants, it may not be the best time period to study the determinants of those grants. There are data limitations, particularly on state and local government finances. The period is short, so several key variables do no vary over the sample, such as Key Pittman's tenure as President Pro Tem of the Senate. Finally, the kind of Congressional forces that Anderson and Tollison suggest should affect grants are more likely to do so over longer periods of time. Congress affects grant allocations through changes in the rules and formulas that are used to allocate grants, while the Executive branch is able to directly and immediately affect the allocation of grants through discretionary provisions in the grant legislation. By examining grant allocations over a longer period of time and using more detailed political control variables, both the congressional and presidential effects on grants become clearer. It is also possible to delineate the effects of state and local expenditures on federal grants.

A new data set was constructed around the information available on federal government grants to states measured in real dollars per capita, with 1967 as the base year. Data on state and local expenditures were taken from the Census of Governments in 1932, 1942, 1962, 1972, and 1982, the 5 years in the panel. Unfortunately, the Census was not conducted in 1952, and there are no comprehensive numbers on local government expenditures in that year. The political productivity measures were constructed for each year based on the preceding ten elections. The congressional variables were constructed in a similar manner, only tenure is counted in years rather than months, from the opening date of the relevant session of Congress. Farm value is measured as the real value, per farm, of land and buildings. Farm population was dropped in favor of the percentage of population living in urban areas. The percentage of land in a state belonging to the federal government was dropped because it has very little variation over the century. The percentage of the population that was white was included from the Reading specification.

Anderson and Tollison noted the problems with collinearity between tenure of the House delegation and the tenure on the House Appropriations Committee. Since control of the House rested with the Democratic party in every year in the sample, I constructed an alternative measure of Appropriations Committee tenure

Table 8.3 Variable means and standard deviations, 1932 to 1982

Variable	Mean	St. dev.
Per capita grants	72.81	64.82
Political productivity	0.00012	0.00010
Standard deviation	0.12	0.05
House tenure	89.95	96.38
Senate tenure	18.42	11.03
House democrat appropriations tenure	7.21	10.22
House republican appropriations tenure	4.13	7.65
Senate democrat appropriations tenure	4.05	7.84
Senate republican appropriations tenure	2.11	5.28
House speaker	0.02	0.14
House majority leader	0.02	0.14
President pro tem	0.02	0.14
Senate majority leader	0.02	0.14
Previous winning vote	0.43	0.11
Average winning vote	0.50	0.11
Farm value (000'S $)	17.45	202.60
Percent urban	0.56	0.20
Per capita income	2436.82	1140.93
Percent white	0.89	0.11
Real per capita local expenditures	257.63	165.54
Real per capita state expenditures	191.40	133.05

which avoids the collinearity problem and produces some interesting results. I counted the tenure of the Democratic members separately from the Republican members, producing four appropriations committee variables, one for the tenure of each party's delegations in the House and Senate committees. House and Senate leadership positions were also entered separately, so there are also four leadership variables.

Means and standard deviations for the variables are found in Table 8.3. Regression results are found in Table 8.4, where the dependent variable in every regression is real per capita federal grants. The first column of Table 8.4 begins with a simple pooled analysis. This regression is similar to the final specification from columns (4) and (5) of Table 8.2. The results, however, are substantially different. Presidential variables appear to have no statistical impact on grants. The overall tenure of a state's Senate delegation and the tenure of the state's Representatives on the House Appropriations committee is positive and significant. But the overall tenure of a state's House delegation exerts a negative impact on per capita grants. Further states with higher incomes receive significantly larger grants and more urban states receive significantly smaller grants.

As mentioned earlier, the political variables used so far are basically swing variables. They measure how close a state's vote is to 50% Democratic, how variable

its voting pattern is, or whether a state's congressmen are in key decision-making positions. But what about rewarding loyal voters? Column (2) of the table includes two political variables that measure voter support for the current President and the party of the current President. "Previous Winning Presidential Vote" is the share of the presidential vote that went to the winning candidate in the previous election. For example, in 1932 the variable measures Hoover's percentage of the popular vote in the 1928 election, while in 1942 the variable measures FDR's percentage of the popular vote in the 1940 election. The variable "Average Winning Presidential Vote" measures the average vote for the party of the current president in the previous ten elections. So the 1932 variable measures the average Republican share in the popular vote from 1892 to 1928, while the 1942 variable measures the average Democratic share of the vote from 1904 to 1940. The sign on the previous Presidential vote is negative, indicating that President's do not reward their supporters, although this conclusion is sensitive to specification. Long term voting patterns appear to have little effect on grant allocation.[15]

Most of the results in column (2), however, are sensitive to changes in specification. The first change is adding state and local expenditures, treating each as an endogenous variable. A set of demographic and fiscal variables are used as instruments for state and local expenditures in the two-stage estimates.[16] The primary effect of including state and local expenditures is to alter the effect of income on grants. The coefficient on per capita income falls from 0.05 with a t-statistic of 0.8. States with higher incomes spend more money and, as a result, receive higher grants. The perverse effect of income on grants appears to be due to this simultaneity, not to the fact that grants programs are designed to give more money to higher income states.

While including state and local expenditures has minor effects on other variables, the inclusion of fixed time and state effects dramatically changes the estimates. Time and state effects are included in column (4). The coefficients on the time and state dummies are not included in the table. F-statistics on the test restricting the fixed effects at zero are presented in the last row of the table. The fixed effects are jointly significant at the 1% levels.

Comparison of column (4) with column (3) of the table shows how important the fixed effects are. The Presidential political productivity coefficient doubles and

[15]This result is in a regression including the Political Productivity variable, which is based on long term voting patterns. Several different alternative ways of including recent voting in Presidential elections were tried. Many were significant in simple panels, but using 2SLS and including fixed effects always eliminated their significance and reduced the magnitude of their coefficients substantially.

[16]The demographic variables are population density, percentage white, percentage male, percentage urban, population, the growth rate of total population int he preceding decade, percent of the population that is school age, and the average size of place of resident of typical urban residents. The economic and fiscal variables are real per capita income, real per capita state debt, real per capita local debt, real per capita assessed value of property, the rank of the state in per capita income tax collections, and the number of governments per capita in the state.

Table 8.4 Regression results, 1932–1982

	OLS (1)	OLS (2)	2SLS (NO FE) (3)	2SLS (FE) (4)
Political productivity	19042	19791	16844	36151**
Standard deviation	11.89	−4.97	70.64**	32.09
House tenure	−0.08**	−0.08	0.00	−0.04
Senate tenure	0.83**	0.75**	0.24	0.14
House dem appropriations tenure	0.87**	0.72**	0.47**	0.50**
House rep appropriations tenure	−0.07	−0.09	−0.35	−0.44**
Senate dem appropriations tenure	0.28	0.27	0.13	0.09
Senate rep appropriations tenure	−0.04	−0.04	−0.29	−0.50*
House speaker	11.56	17.49	7.24	10.12
House majority leader	5.12	4.96	2.54	2.38
Senate majority leader	24.45	23.46	23.24**	17.68**
President pro tem	−8.43	−6.78	−2.18	6.91
Previous winning vote		−0.60**	−0.33**	−0.04
Average winning vote		0.27	0.13	0.03
Farm value (000'S $)	0.004	0.004	−0.0123*	−0.002
Percent urban	−81.42**	−86.09**	−26.82**	41.23**
Per capita income	0.053**	0.053**	−0.003	−0.007
Percent white	−8.07	−9.91	−1.48	−50.10
Real per capita local expenditures			0.08**	−0.05
Real per capita state expenditures			0.41**	0.31**
Constant	−21.57	5.52	−8.74	19.98
R−squared	0.73	0.73		
F	37.3	33.9	111.6	53.9
SSE	273232	267026	89133	42096
F-statistic on FE prob				1.74*

Note: Dependent variable in each regression is real per capita state grants. T-statistics not included for space

*10% significance; *5% significance level

becomes significant. The standard deviation of the democratic share of the vote becomes insignificant. The percentage of a state's population that is urban now exerts a positive and significant effect on grants instead of a negative effect, while income exerts a negative but not significant effect. State government expenditures remain an important positive determinant of grants, and local expenditures become statistically insignificant.

The results in Table 8.4 indicate the specification dangers inherent in these estimates. Major results are reversed when more appropriate specifications are adopted. Using the coefficient estimates in column (4) as a baseline, we can draw some conclusions regarding the relative importance of political, economic, and fiscal factors in the allocation of grants.

Wright's presidential productivity index and Democratic tenure on the House Appropriations Committee are significant political variables, both statistically and quantitatively. A one standard deviation increase in the political productivity index produces a $3.66 change in grants per capita per year, a 5% shift in the mean annual grant of $72. Increasing a state's Democratic tenure on the Appropriations committee by a standard deviation of 10.22 years produces a $5 increase in grants, a shift of 7%. Increasing the tenure of a state's Senate delegation always increases grants, though significance is specification sensitive, and an increase of eleven years produces an increase of $1.54, a shift of only 2%. Republican tenure on both Appropriations committees, however, is consistently negative and significant when fixed effects are controlled for. A one standard deviation in Republican committee tenure reduces grants by $3.37 in the House and $2.64 in the Senate.

The Senate Majority Leader seems to exert the most important affect on grant allocations. When a state's Senator becomes majority leader, per capita grants rise by $17.68, an increase of 25% over the mean annual grant. It is more difficult to evaluate the overall impact of this variable, since a one standard deviation increase in the Senate majority leader would produce an increase in grants of only $2.47. It is, of course, impossible to have a 0.14 increase in the majority leader.

The quantitative impact of the key economic variables is larger than any of the political variables except for the Senate majority leader (although the income and local government coefficients are statistically insignificant). Increasing percent urban by a standard deviation raises grants by $8.27 or 11%, and increasing income by a standard deviation lowers grants by $7.99 or 11%. The fiscal variables are quite important. Increasing local expenditures by a standard deviation lowers grants by $8.28 or 11%. Raising state expenditures by a standard deviation raises grants by $41.24 or 57% of the mean annual grant. The much larger effect of state expenditures on grants is due entirely to the estimated coefficients, the standard deviation of local expenditures is actually higher than state expenditures, $165 to $133. Variations in state expenditures are far and away the most influential determinant of federal grants.

Like the New Deal, the twentieth century experience with grants has been one where economic and political factors have an impact on the allocation of federal grants to the states. Unlike the New Deal, it is difficult to make a case that Presidential politics were more important that congressional politics, both seem to be important. Quantitatively the most important congressional variables were tenure on the House Appropriations Committee and Senate Majority Leader. Tenure itself, however, is not necessarily rewarded, as shown by the negative effect of Republican tenure on the Appropriations committees. Over the long term it seems that economic and fiscal factors exerted a larger effect on the allocation of grants than political factors. The economic effects work in the right way. The single most important determinant of federal grants is state expenditures. This is consistent with the way grants actually work in the United States and a more complicated picture of the process underlying the allocation of grants.

8.5 Conclusions

This paper has three conclusions. The first is econometric. When trying to disentangle the effect of economic factors and presidential and congressional politics on the allocation of federal largess to the states, it is important to keep in mind that (1) economic and political variables may affect state as well as federal behavior so that a model with endogenous grants is called for, and (2) since we know little about the specific politics of grant allocation, we should try to control for "unobservable" state and time effects. The results presented here show how important it is to control for both of these problems. Most of the key coefficients in the twentieth century analysis are affected by changes in specification. The effects of income and urbanization on grants appear to be perverse in a simple pooled-cross sectional regression, but inclusion of fixed effects and accounting for simultaneous grants and state and local expenditures solves the problem.

Second, states matter. The structure of government in the United States in this century has moved progressively toward independent federalism. In this structure, state governments exert independent decision-making power, and their decisions have an obvious impact on the federal government. While it is common to think of the federal government as calling the tune in its relationships with the states, the reality is far more complicated.

The third conclusion is that all of the major influences on federal allocations to the states identified in this debate appear to have some effect on national policy. During the New Deal economic factors did influence spending in the way New Dealers suggested it should have. And the positive effect of state income on federal allocations may, someday, be resolved by better information on state and local government expenditures, as is strongly suggested by the longer panel data set. Presidential politics, as measured by Wright's variables, played an important role. Whether this was purely presidential politics or whether it was what we might call national politics, isn't clear. Democrats in the House and Senate may have had a vested interest in keeping a Democrat in the White House. Presidents may have used their discretionary powers to help strategically placed Senators and Representatives. The pattern of allocation, however, in both the New Dean and longer sample suggests that national electoral outcomes do play a role in determining grant allocations.

Congressional factors appear to be considerably more important over the long run than they were in the New Deal. The tenure of the Democrats on the House Appropriations Committee and Senate Majority Leadership exert a significant effect on grants. It is difficult to make a case, however, that either presidential or congressional influences mattered more over the longer period.

There is, however, clear support for the argument that presidential politics mattered more during the New Deal. In the panel data set for the 1930s, Wright's political variables continue to explain a good portion of the variability in grants across states, while the congressional variables all become insignificant. This is particularly important when the peculiarities of Nevada are taken into account.

The importance of presidential factors is consistent with the flexibility given to Roosevelt in the New Deal programs. That flexibility was gradually taken away from Presidents, even from Roosevelt after 1937.

That presidential, congressional, economic, and fiscal factors influenced the allocation of federal grants should come as no surprise to a student of the history of fiscal structure in the twentieth century. Indeed, it would be surprising to find that only one factor was important. Intergovernmental grants have become a major part of the fiscal structure. They exit for social and economic reasons that reflect the interests of the electorate, for public finance reasons that reflect the benefits from centralized collection of revenues and decentralized administration of expenditures, as well as political reasons. Grants represent tangible benefits that can be delivered by politicians. For all these reasons grants will continue to be used for a long time, and their determinants will reflect the complex interaction of economic, fiscal, and political forces.

Acknowledgements Some of the data in this paper are the result of the generosity of Gavin Wright, and the support of George Stigler when I was a fellow at the Center for the Study of the Economy and the State. Farley Grubb was my research assistant then, and I have more recently been helped by Robert Satterfield and Brad Lewis. Wally Oates, Shawn Kantor, Price Fishback, and the Development of the American Economy program meeting participants provided helpful comments.

References

Advisory Council of Intergovernmental Relations (1994) Significant features of fiscal federalism. Advisory Council of Intergovernmental Relations, Washington DC

Anderson GM, Tollison RD (1991) Congressional influence and patterns of New Deal spending, 1933–1939. J Law Econ 34(1):161–175

Arrington L (1969) The New Deal in the west: a preliminary statistical inquiry. Pac Hist Rev 38(3):311–316

Arrington LJ (1970) Western agriculture and the New Deal. Agric Hist 44(4):337–353

Benson G (1941) The new centralization: a study of intergovernmental relationships in the United States. Farrar & Rinehart, New York

Bittermann HJ (1938) State and federal grants-in-aid. Mentzer, Bush & Company, Chicago

Bungey M, Grossman P, Kenyon P (1991) Explaining intergovernmental grants: Australian evidence. Appl Econ 23(4):659–668

Chernick HA (1979) An economic model of the distribution of project grants. In: Mieszkowski PM, Oakland WH (eds) Fiscal federalism and grants-in-aid. Urban Institute, Washington DC, pp 81–103

Clark JP (1938) The rise of a new federalism: federal-state cooperation in the United States. Russell & Russell, Washington DC

Craig SG, Inman RP (1982) Federal aid and public education: an empirical look at the new fiscal federalism. Rev Econ Stat 64(4):541–552

Gramlich EM (1977) Intergovernmental grants: a review of the empirical literature. In: Oates WE (ed) The political economy of fiscal federalism. Lexington Books, Lexington, pp 219–239

Holcombe RG, Zardkoohi A (1981) The determinants of federal grants. South Econ J 48(2): 393–399

Howard DS (1943) WPA and federal relief policy. Russell Sage Foundation, New York

Ladd HF (1991) The state aid decision: changes in state aid to local governments, 1982–87. Natl Tax J 44(4):477–496
Oates WE (1972) Fiscal federalism. Harcourt Brace Jovanovich, New York
Office of Government Reports, Statistical Section (1940) Report no. 10, vol II. Government Printing Office, Washington DC
Reading DC (1973) New Deal activity and the states, 1933 to 1939. J Econ Hist 33(4):792–810
Wallis JJ (1984) The birth of the old federalism: financing the New Deal, 1932–1940. J Econ Hist 44(1):139–159
Wallis JJ (1987) Employment, politics, and economic recovery during the Great Depression. Rev Econ Stat 69(3):516–520
Wallis JJ (1989) Employment in the Great Depression: new data and hypotheses. Explor Econ Hist 26(1):45–72
Wallis JJ (1991) The political economy of New Deal fiscal federalism. Econ Inq 29(3):510–524
Wiliams EA (1939) Federal aid for relief. Columbia University Press, New York
Wright G (1974) The political economy of New Deal spending: an econometric analysis. Rev Econ Stat 56(1):30–38
Yinger J, Ladd HF (1989) The determinants of state assistance to central cities. Natl Tax J 42(4):413–428