

Does Population Heterogeneity Really Matter to Nonprofit Sector Size? Revisiting Weisbrod's Demand Heterogeneity Hypothesis

Jiahuan Lu¹ 

Published online: 10 November 2017

© International Society for Third-Sector Research and The Johns Hopkins University 2017

Abstract The size variation of the nonprofit sector across geographic locations has attracted substantial scholarly attention. One dominant theory exploring this issue is Weisbrod's demand heterogeneity hypothesis which suggests that the nonprofit sector size in a locality is positively related to population heterogeneity of the area. However, empirical tests of Weisbrod's hypothesis reach a very inconsistent conclusion. This study employs a meta-analysis to aggregate existing studies on the relationship between population heterogeneity and nonprofit sector size. The study finds a significant and positive association between the two variables, but the magnitude of the relationship is substantially small. Further, this relationship seems generalized across countries and is supported by both within-country and cross-country data and by different measurements of nonprofit sector size. In addition, population heterogeneity in terms of age, education, ethnicity, language, and religion supports the hypothesis better. Overall, although the demand heterogeneity hypothesis holds empirically, its explanatory power in predicting nonprofit sector size and growth might be less robust.

Keywords Government failure · Demand heterogeneity · Population heterogeneity · Nonprofit sector size

Résumé Les différences de taille du secteur sans but lucratif de différents lieux géographiques ont attiré l'attention de nombreux savants. Une des principales théories

explorant cet enjeu est l'hypothèse d'hétérogénéité de la demande de Weisbrod, qui suggère que la taille du secteur sans but lucratif d'une région est positivement associée à l'hétérogénéité de sa population. Les tests empiriques auxquels l'hypothèse de Weisbrod a été soumise procurent cependant des résultats fortement incohérents. La présente étude regroupe de nombreuses études existantes sur la relation entre l'hétérogénéité de la population et la taille du secteur sans but lucratif. Elle démontre qu'il existe une association positive et importante entre les deux variables, mais que la relation est essentiellement minime. Qui plus est, la relation semble généralisée à travers les pays et soutenue par des données nationales et internationales, ainsi que par différentes mesures de la taille du secteur sans but lucratif. En outre, l'hétérogénéité de la population, en termes d'âge, d'éducation, d'ethnicité, de langue et de religion, soutient l'hypothèse davantage. En fin de compte, même si l'hypothèse de l'hétérogénéité de la demande tient la route d'un point de vue empirique, sa capacité à expliquer la taille du secteur sans but lucratif peut être moins robuste.

Zusammenfassung Die Größenvariation des gemeinnützigen Sektors je nach geographischem Standort erhält große Aufmerksamkeit von Wissenschaftlern. Eine dominante Theorie, die sich mit diesem Thema beschäftigt, ist Weisbrods Hypothese der Nachfrageheterogenität, derzufolge die Größe des gemeinnützigen Sektors an einem Ort positiv mit der Bevölkerungsheterogenität in der Region korreliert. Allerdings kommen empirische Tests der Hypothese von Weisbrod zu einer äußerst inkonsistenten Schlussfolgerung. Diese Studie wendet eine Metaanalyse an, um bestehende Studien zur Beziehung zwischen der Bevölkerungsheterogenität und der Größe des gemeinnützigen Sektors zusammenzufassen. Das Ergebnis

✉ Jiahuan Lu
jjiahuan.lu@rutgers.edu

¹ School of Public Affairs and Administration, Rutgers University-Newark, 111 Washington Street, Newark, NJ 07102, USA

der Studie zeigt eine signifikante und positive Verbindung zwischen den beiden Variablen, wobei das Ausmaß der Beziehung jedoch sehr gering ist. Darüber hinaus scheint diese Beziehung in unterschiedlichen Ländern generalisiert zu sein und wird sowohl von inländischen als auch länderübergreifenden Daten sowie von verschiedenen Größenmessungen des gemeinnützigen Sektors gestützt. Des Weiteren wird die Hypothese besser von der Bevölkerungsheterogenität in Bezug auf Alter, Bildung, Ethnizität, Sprache und Religion gestützt. Insgesamt hält die Hypothese der Nachfrageheterogenität empirisch stand, ihre Erklärungskraft zur Voraussage der Größe und des Wachstums des gemeinnützigen Sektors ist hingegen unter Umständen weniger robust.

Resumen La variación del tamaño del sector de las organizaciones sin ánimo de lucro en zonas geográficas ha atraído una importante atención por parte de los eruditos. Una teoría dominante que explora esta cuestión es la hipótesis de la heterogeneidad de la demanda de Weisbrod que sugiere que el tamaño del sector de las organizaciones sin ánimo de lucro está relacionado de manera positiva con la heterogeneidad de la población de la zona. Sin embargo, pruebas empíricas de la hipótesis de Weisbrod llegan a una conclusión muy contradictoria. El presente estudio emplea un meta-análisis para acumular estudios existentes sobre la relación entre la heterogeneidad de la población y el tamaño del sector de las organizaciones sin ánimo de lucro. El estudio encuentra una asociación positiva y significativa entre las dos variables, pero la magnitud de la relación es sustancialmente pequeña. Asimismo, esta relación parece generalizada en todos los países y se ve soportada tanto por datos dentro del país como por datos entre los países y por diferentes mediciones del tamaño del sector de las organizaciones sin ánimo de lucro. Igualmente, la heterogeneidad de la población en términos de edad, educación, etnicidad, idioma y religión apoya mejor la hipótesis. En general, aunque la hipótesis de la heterogeneidad de la demanda se mantiene empíricamente, su poder explicativo para predecir el tamaño y el crecimiento del sector de las organizaciones sin ánimo de lucro podría ser menos sólido.

Introduction

The nonprofit sector has been increasingly acknowledged by scholars and practitioners as a vital and indispensable component of a nation's economic and social affairs. The nonprofit community serves fundamental social needs, such as caring for the most vulnerable people, educating youth and adults, organizing for social change, and providing spiritual fulfillment. In addition, the nonprofit sector

is often the source of innovative solutions to pressing public problems and a major player in how public policy is developed and implemented. Over the past several decades, the nonprofit sector has grown in size, sophistication, and impact in various countries all over the world, including Australia (Lyons 1993), Brazil (Marchesini da Costa 2016), Japan (Kanaya et al. 2015), Netherlands (Burger and Veldheer 2001), South Korea (Kim and Kim 2015), Spain (Marcuello 1998), and USA (McKeever 2015).

Despite steady development, the growth pattern of the nonprofit sector and the distribution of nonprofit activities differ significantly across communities, localities, and countries. For example, in the USA, the northeastern part of the country has higher nonprofit density, while the South and the Rocky Mountain areas have a relatively smaller nonprofit sector (Liu 2016). Internationally, nonprofit sector employment (in terms of paid and volunteer workers as a percentage of total national workforce) across countries also varies from a low of 0.9% to a high of 12.7% in Israel, with countries such as Brazil, Kyrgyzstan, and Czech Republic grouped toward the low end of the spectrum and Australia, Belgium, and New Zealand toward the higher end (Salamon et al. 2013). Indeed, this variation of nonprofit sector size across geographic locations has been documented by many studies, including Bielefeld and Murdoch's (2004) study of education and human service nonprofits in U.S. metropolitan areas, Grønbjerg and Paarlberg's (2001) study of Indiana nonprofit sector, James's (1987) international comparison of educational nonprofit sector, Kim and Kim's (2016b) examination of OECD countries, Marchesini da Costa's (2016) study of nonprofits across Brazilian municipalities, Marcuello's (1998) study of nonprofits in Catalonia, Spain, and Oktan and Osili's (2004) study of community organizations in Indonesia, just to name a few.

To account for the variation in nonprofit activities and growth pattern across geographic locations, over the past several decades a number of theories on nonprofit formation and behaviors have been proposed from diverse disciplines and perspectives (e.g., Ben-Ner and Van Hoomissen 1991; DiMaggio and Anheier 1990; Hansmann 1980; Lecy and Van Slyke 2013; Rose-Ackerman 1996; Salamon 1987; Salamon and Anheier 1998; Steinberg 2006; Weisbrod 1988). This body of theories greatly contributes to our understanding of the nature and roles of nonprofits in society. Among them, one of the most influential theories is the government failure theory developed by Economist Weisbrod (1975) to explain why nonprofits exist in a mixed-sector economy.¹ The basic model of the

¹ Some scholars also term Weisbrod's government failure theory the public goods theory, while others use the two names interchangeably (e.g., Kingma 2003, Salamon and Anheier 1998, Steinberg 2006). The present paper calls the theory the government failure theory.

theory, the demand heterogeneity hypothesis, argues that in heterogeneous communities where citizens' tastes for public goods are more diverse than what median voters prefer, there will be more nonprofits established to provide public goods to satisfy the demand unmet by government provision. The level of nonprofit activities (hereafter referred to as "nonprofit sector size") in a geographic location is thus determined by the population heterogeneity of the location. The more diverse the population in a community is, the larger the nonprofit sector we could expect.

Since its inception, Weisbrod's demand heterogeneity hypothesis has been widely tested and extended (e.g., Corbin 1999; DiMaggio and Anheier 1990; Kingma 2003; Lecy and Van Slyke 2013; Steinberg 2006). Interestingly, this body of literature reaches mixed conclusions about the relationship between population heterogeneity and nonprofit sector size. Although some studies find the expected positive association between the two variables (e.g., Corbin 1999; James 1987; Joassart-Marcelli 2013; Lu 2016; Matsunaga et al. 2010; Wolch and Geiger 1983), many other studies report null or even negative association (e.g., Abzug and Turnheim 1998; Grønberg and Paarlberg 2001; Kim and Kim 2016a; Marchesini da Costa 2016; Marcuello 1998; Salamon and Anheier 1998; Van Puyvelde and Brown 2016). These inconsistent findings, although they provide a rich examination of the demand heterogeneity hypothesis from different lenses, cause difficulty in providing a coherent assessment of the demand heterogeneity hypothesis and a consistent knowledge base to guide future nonprofit studies. There is a scholarly need to make sense of this body of existing studies and draw a generalized conclusion from competing findings.

The present study employs a meta-analysis technique to aggregate existing quantitative tests of the demand heterogeneity hypothesis. Through a quantitative systematic review of 37 existing studies with a total of 491 effect sizes, we find a significant and positive association (a correlation coefficient r between .020 and .147) between population heterogeneity and nonprofit sector size in a geographic location. Population heterogeneity tends to be positively related to nonprofit formation and growth—basically in line with the demand heterogeneity hypothesis—but the magnitude of the relationship seems to be substantially small. Further, this relationship seems to be generalized across countries. On the research design side, the demand heterogeneity hypothesis is supported by both within-country and cross-country data and by different measurements of nonprofit sector size. In addition, some dimensions of population heterogeneity (including age, education, ethnicity, language, and religion) seem to support the demand heterogeneity hypothesis better than other dimensions (including employment status, gender,

household location, income, and origin). Overall, although the demand heterogeneity hypothesis holds empirically, its explanatory power in predicting nonprofit sector size and growth might be less robust. Future studies might be cautious when applying the demand heterogeneity hypothesis to explain the size variation of the nonprofit sector.

The remaining sections are arranged as follows. The second section reviews Weisbrod's government failure theory and its demand heterogeneity hypothesis in particular. Next, we detail the procedures of the meta-analysis in section three. Section four presents the results of the meta-analysis. Finally, we conclude the paper in section five with a discussion of the implications of the findings as well as the limitations of the meta-analysis and existing studies.

Weisbrod's Demand Heterogeneity Hypothesis Revisited

Weisbrod's (1975, 1977a, 1986, 1988) government failure theory has been widely acknowledged as a cornerstone in the literature on the origins of nonprofit organizations. The key question that Weisbrod attempted to answer is why nonprofits exist in market economies. Holding an economic lens, Weisbrod started with the notion of government as a response to market failure in providing public goods because of problems such as free riders. He explained that government can also fail to provide sufficient public goods to meet citizen needs in all circumstances, largely because of the tension between demand heterogeneity and median voters. As a society grows more and more demographically diverse, the diversified population would ask for diverse public goods, in terms of quality, style, level, etc. As such, citizens would expect government to be more responsive to the needs of these diverse groups and ideally tailor public goods to their diversified demands.

However, this request creates a dilemma for government. On the one hand, government cannot respond to the preferences of all the constituent groups in local communities. Often, these diversified goods are only desired in accordance with local preferences and consumed in local communities. Unless a particular set of collective goods is demanded by a sizable political constituency, these local collective demands might not be large enough to be addressed by government intervention as a public policy issue. The government's bureaucratic process and accountability environment would also limit government response to citizen needs (Wilson 2000). On the other hand, probably more fundamentally, the quality and quantity of government provision in a democratic society is decided by a majority of voters through the political voting process. As a result, the majority tends to choose the level of government provision that median voters prefer. Public

officials and political leaders, to maximize their chances of re-election, will ensure that public service provision is at the level that most voters prefer, leaving other service needs unsatisfied in the public choice process (Buchanan and Tullock 1962). In this way, “when demand is diverse, though, whatever quantities and qualities of services government provides will oversatisfy some people and undersatisfy others” (Weisbrod 1988, p. 25), resulting in government failure in meeting the needs of all citizens.

Under this circumstance, Weisbrod argued that nonprofit organizations, supported by charitable donations and voluntary activities from individuals who share the same local public interests, can function as “extragovernmental providers of collective-consumption goods” (Weisbrod 1977a, p. 59) to meet the residual and particularistic service needs.² Indeed, several nonprofits’ defining features might better position them to act as a supplement to government provision. As grassroots community-based organizations, nonprofits are embedded in local communities to varying degrees and thus have better understanding of local needs. Additionally, compared to government, nonprofit operations are subject to less bureaucratic control and political oversight and thus could be more flexible and efficient in responding to social and community issues (Douglas 1987). In sum, nonprofits seem to be more readily to tailor goods to local diversified needs. In this way, the existence of nonprofits can be justified as “an alternative to the private-sector provision of the private-good substitutes for collective goods” (Weisbrod 1977a, pp. 59–60).

Since Weisbrod’s seminal work in the 1970s, the government failure theory has been considered by scholars as a classic on the economics of nonprofit organizations. A major test of this theory is a test of the demand heterogeneity hypothesis (Kingma 2003; Matsunaga et al. 2010).³ As Matsunaga et al. (2010, p. 182) claimed, in Weisbrod’s theory “the core source of government failure is demand heterogeneity, and it is the most influential factor determining the size of the nonprofit sector.” The hypothesis, in Weisbrod (1988, p. 27) words, states “if government responds to the demands of the majority and the nonprofit

sector responds to the demands of the undersatisfied, then the greater the diversity of demand the larger the size of the nonprofit sector will be, other things being equal.” Given the difficulty in measuring demand heterogeneity, Weisbrod further suggested that demand heterogeneity can be proxied by population heterogeneity in terms of income, education, age, ethnicity, etc. He (1986, p. 37) noted that “the degree of heterogeneity in quantities demanded is positively correlated with the degree of heterogeneity in preferences, and this, in turn, may be proxied by the degree of heterogeneity of the population in the given governmental unit”. In this way, the demand heterogeneity hypothesis becomes: the size of the nonprofit sector in a geographic region is positively related to the population heterogeneity of that region.

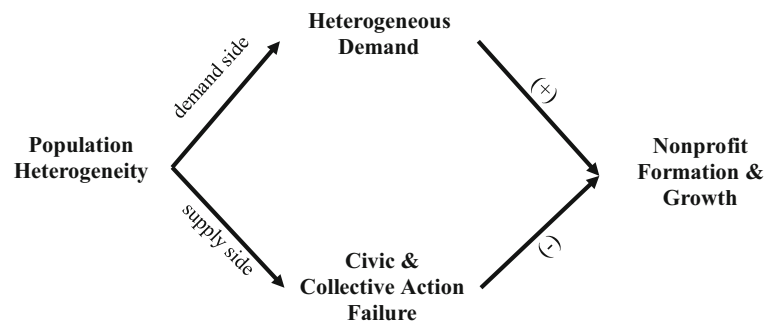
Indeed, since its inception, the demand heterogeneity hypothesis has been extensively examined and extended (e.g., Corbin 1999; DiMaggio and Anheier 1990; Kingma 2003; James 1987; Lecy and Van Slyke 2013; Liu 2016; Matsunaga and Yamauchi 2004; Steinberg 2006). A major critique of Weisbrod’s theory is its focus on the demand side of nonprofit formation but its ignorance of the supply side (Ben-Ner and Van Hoomissen 1991; DiMaggio and Anheier 1990; James 1987). Indeed, the demand heterogeneity hypothesis seeks to explain the existence of nonprofits as a function of heterogeneous demand brought by population heterogeneity, but it neglects consideration of demand-side stakeholders or entrepreneurs such as consumers, sponsors, and donors who ultimately establish or support nonprofits to respond to residual demands. On its own, the existence of demand for nonprofits does not necessarily guarantee a supply. Rather, as Ben-Ner and Van Hoomissen (1991, p. 540) forcefully argued, “there must be a confluence between the demand for the organizational form and the ability to provide it in order for a nonprofit organization to be formed.” In this way, although the process of nonprofit formation can be demanded by population heterogeneity, nonprofits will not materialize until stakeholders exist to generate the supply. Therefore, without readily exploring the supply-side effect of population heterogeneity on nonprofit formation, the explanatory power of the demand heterogeneity hypothesis could be greatly undermined.

In fact, a growing body of recent literature documents that population heterogeneity can be detrimental to civic engagement and collective actions. In demographically heterogeneous communities, people display a lower level of trust and social interactions (Alesina and La Ferrara 2002; Costa and Kahn 2003; Putnam 2007), get involved in fewer social and civic activities (Alesina and La Ferrara 2000; Anderson and Paskeviciute 2006; Costa and Kahn 2003), and contribute less to community organizations and public fundraising (Andreoni et al. 2016; Fong and Luttmer

² Weisbrod (1977a) also discussed other options for individuals to address unmet demands, such as moving to a different political unit, forming a lower-level government unit to provide additional service, and substituting private goods for collective goods.

³ According to Lee and Weisbrod (1977, p. 77), the government failure theory includes three hypotheses, the output hypothesis (the nonprofit sector produces collective goods), the income hypothesis (the demand for collective goods has an inverted *U*-shape relationship with income), and the demand heterogeneity hypothesis (the importance of nonprofits in providing collective goods has a positive relationship with demand heterogeneity). The present paper only tests the demand heterogeneity hypothesis. Literature on the other two hypotheses can be found in such studies as Kingma (2003) and Salamon and Anheier (1998).

Fig. 1 The impact of population heterogeneity on nonprofit formation and growth



Note: (+) represents positive effect and (-) represents negative effect.

2009; Miguel and Gugerty 2005; Okten and Osili 2004). There are two possible reasons for this negative effect. First, people are inclined to trust and associate with others who are demographically similar to themselves because of shared norms, interests, and empathy (Alesina and La Ferrara 2000, 2002; Glaeser et al. 2000). Second, diverse groups have divergent preferences or agendas, which makes it extremely difficult to bridge competing interests and take collective actions toward a shared goal (Habyarimana et al. 2007; Miguel and Gugerty 2005). Alesina and La Ferrara (2000) employed U.S. locality data and found that population heterogeneity (in terms of income and ethnicity) significantly reduces participation in community activities such as religious, civic, educational, and service groups. Putnam (2007) studied social capital across U.S. communities and concluded that in communities with greater diversity, people are less likely to trust their neighbors, collaborate on community projects, and donate their time and money. In sum, population heterogeneity might limit the supply of social capital and other voluntary resources that are fundamental antecedents to nonprofit formation and development (Ben-Ner and Van Hoomissen 1991; Saxton and Benson 2005).

Therefore, in the consideration of the demand heterogeneity hypothesis, when population heterogeneity leads to diversified needs that call for more nonprofits to fill the gap, it simultaneously undermines the social capital and charitable activities that sponsor nonprofit formation. In this way, the impact of population heterogeneity on nonprofit formation and growth depends on the interaction between its demand-side and supply-side effects (see Fig. 1). If the demand-side and supply-side effects balance each other, the relationship between population heterogeneity and nonprofit sector size might become ambiguous.

Indeed, existing empirical tests of the demand heterogeneity hypothesis reach a mixed conclusion concerning the relationship between population heterogeneity and nonprofit sector size. Some studies find a positive relationship between the two and thus endorse the demand heterogeneity hypothesis (e.g., Corbin 1999; James 1987;

Joassart-Marcelli 2013; Lu 2016; Matsunaga et al. 2010; Wolch and Geiger 1983). However, some studies report that population heterogeneity has almost no significant impact on nonprofit sector size (e.g., Abzug and Turnheim 1998; Garrow and Garrow 2014; Salamon and Anheier 1998; Salamon et al. 2000). Still other studies conclude that population heterogeneity actually has a negative association with nonprofit existence and growth (e.g., Grønberg and Paarlberg 2001; Kanaya et al. 2015; Kim and Kim 2016a; Marchesini da Costa 2016; Marcuello 1998; Van Puyvelde and Brown 2016)—just the opposite to the demand heterogeneity hypothesis. Considering all the empirical evidence, there seems to be no consensus as to whether the population hypothesis matters to nonprofit sector size and if so in what way.

In sum, although Weisbrod's demand heterogeneity hypothesis has been widely considered as a classic model on nonprofit formation and behaviors, inconsistent findings from existing empirical studies question the explanatory power of the hypothesis and present a significant barrier to forming a coherent knowledge base to guide future nonprofit research and practice.

Method

We explored the relationship between population heterogeneity and nonprofit sector size by employing a meta-analysis technique. Given that the demand heterogeneity hypothesis has been empirically tested extensively over the past four decades, a more critical scholarly need might be not to add additional empirical studies but to take stock of these existing studies and integrate conflicting findings to establish a generalized understanding. This is where meta-analysis could help. Meta-analysis, as Glass (1976, p. 3) coined it, represents “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.” As a quantitative research synthesis technique, meta-analysis enables researchers to draw statistical information from existing

original studies and then aggregate it to form a more accurate conclusion across different studies. In this way, meta-analysis can build bodies of cumulative knowledge and provide a more precise estimation of the underlying relationship between population heterogeneity and nonprofit sector size, which further contributes to the development of government failure theory and its demand heterogeneity hypothesis in particular.

Literature Search

As a systemic review method, meta-analysis starts with a literature search to identify all the existing studies concerning the impact of population heterogeneity on nonprofit sector size in a geographic area. To promote the comprehensiveness of the literature search, we employed three complementary search strategies in Google Scholar to identify as many relevant studies as possible (Reed and Baxter 2009).⁴ First, given that studies testing Weisbrod's demand heterogeneity hypothesis have to cite Weisbrod's seminal work, we searched all the studies that cite Weisbrod's original paper (Weisbrod 1975), its subsequent reprints (Weisbrod 1977a, 1986) and books (Weisbrod 1977b, 1988).⁵ Second, we undertook an ancestor search to examine the references of the studies we found to seek other relevant studies. Third, we conducted a descendant search to identify later studies that cite the studies identified by the previous two search strategies. We iterated the second and third steps until no new relevant studies could be identified. The entire literature search was concluded on December 16, 2016. In most cases, we manually reviewed the abstracts of all the relevant studies. If an article's eligibility for inclusion was unclear from its abstract, we performed a full-text review to determine whether the study meets our criteria.

Inclusion Criteria

The following inclusion criteria were used to identify primary studies to be included in the present meta-analysis: (1) the study includes at least one quantitative test of the effect of population heterogeneity on nonprofit sector size in a locality, (2) the focal predictor, *population*

heterogeneity, represents one or several dimensions of a population's demographic diversity (e.g., income, age, ethnicity, gender, and education), (3) the dependent variable, *nonprofit sector size*, refers to the size of the nonprofit sector in a given geographic area (e.g., organization number, employment size, and expenditure amount), and (4) the study provides sufficient statistical information to calculate effect sizes. After detailed screening, we finally included 37 original studies for the meta-analysis, with a wide variety of nonprofit types, countries, and research designs represented in these studies (see Table 1), which promotes the external validity of the present analysis.

Coding Procedures

Information from these accepted studies was then extracted and coded. In particular, two types of information were included in the synthesis: effect size information and study characteristics (Lipsey 2009). Effect size, the key metric in meta-analysis, refers to the standardized association between focal predictor and dependent variables. In the present study, effect size represents the standardized association between population heterogeneity and nonprofit sector size. Given that most existing studies measure both population heterogeneity and nonprofit sector size continuously, we employed a correlation-based (*r*-based) effect size (transformed into Fisher's *z* to correct the small bias associated with correlation coefficient *r*) to represent the population heterogeneity—nonprofit sector size relationship in original studies (Borenstein 2009; Ringquist 2013).⁶

When calculating correlation-based effect sizes from original studies, we employed the following coding strategies suggested by Borenstein (2009), Fleiss and Berlin (2009) and Ringquist (2013): (1) When original studies report odds ratios (e.g., Garrow and Garrow 2014), odds-based effect sizes were first coded and then converted into correlation-based effect sizes. (2) When original studies only report parameter estimates and indicate their statistical significance levels using asterisks (e.g., Bielefeld et al. 1997; Bielefeld and Murdoch 2004), *t*-scores or *z*-scores were estimated as the values at the symbol levels and then converted into correlation-based effect sizes. In this situation, the effect sizes represent lower bound estimates. (3) When original studies only report the parameters of interest that are not statistically significant (e.g., Peck 2008; Salamon and Anheier 1998), the effect sizes were set as 0. In this situation, the effect sizes again represent lower

⁴ Google Scholar provides a comprehensive coverage of scholarly literature in a variety of publishing formats such as journal articles, books, book chapters, working papers, conference papers, and dissertations. The reliance on Google Scholar in the literature search thus allows us to reach a diverse set of studies.

⁵ As of November 30, 2016, Weisbrod's seminal paper *Toward a Theory of the Voluntary Nonprofit Sector in a Three-Sector Economy* (including its 1975; 1977a; 1986 versions) has been cited by 1386 studies; Weisbrod's 1977a, b book *The voluntary nonprofit sector: An economic analysis* has been cited by 276 studies; Weisbrod's 1988 book *The nonprofit economy* has been cited by 1952 studies.

⁶ Transformation to Fisher's *z* could help normalize our effect sizes to correct the skewed distribution of Pearson's *r* around a given population, which allows us to estimate average effect sizes with more accuracy. However, for an individual effect size, the difference between a Fisher's *z* and a correlation coefficient *r* is generally smaller than rounding error (Ringquist 2013).

Table 1 Studies included in the meta-analysis

#	Study	Year coverage	Sample size	Data structure	Type of nonprofits	Nonprofit country	Level of analysis	Measurement of nonprofit size	Measurement of population heterogeneity	# of effect size
1	Abzug and Turnheim (1998)	1987–1989	50	Cross-sectional	Public charities	USA (nationwide)	State	Organizational density	Employment, ethnicity	24
2	Bendick (1977)	1970	102	Cross-sectional	Education	USA (nationwide)	State	Market share	Income, education, ethnicity, religion	4
3	Ben-Ner and Van Hoomissen (1991)	1982–1987	112	Cross-sectional	Education, health, social service	USA (New York state)	County	Employment size	Age, education, ethnicity, household location ^b	18
4	Bielefeld and Murdoch (2004)	1999	540	Cross-sectional	Education, human service	USA (metropolitan areas)	Census tract	Financial size	Age, ethnicity, income	24
5	Bielefeld et al. (1997)	1994	1112–1666	Cross-sectional	Health, education, social service	USA (Dallas county, TX)	Census tract	Organizational density	Age, ethnicity, income	45
6	Corbin (1999)	1992	285	Cross-sectional	Social service	USA (nationwide)	Metropolitan area	Organizational density	Ethnicity, religion	4
7	Garrow and Garrow (2014)	2001–2011	2032	Cross-sectional	Human service	USA (Los Angeles county)	Census tract	Organizational density	Employment, ethnicity	12
8	Grønberg and Paarberg (2001)	1999	92	Cross-sectional	Charitable, advocacy, membership	USA (Indiana)	County	Organizational density	Religion	8
9	James (1987)	Not specified	11–50	Cross-sectional	Education	Cross-nation (4 countries)	Country	Market share	Education, ethnicity, religion	8
10	James (1993)	1975	100	Cross-sectional	Education	Cross-nation (40 countries)	Country	Market share	Income, language, religion	12
11	Joassart-Marcelli (2013)	2006	227	Cross-sectional	Immigration	USA (Greater Boston)	Municipality	Organizational density, financial size	Ethnicity, origin	10
12	Joassart-Marcelli and Wolch (2003)	1996	153	Cross-sectional	Anti-poverty	USA (southern California)	City	Organizational density, financial size	Education, employment, ethnicity, origin	26
13	Kanaya et al. (2015)	2000–2007	376	Longitudinal	Long-term care	Japan (nationwide)	County	Market share	Age, education	6
14	Kim (2015)	2008	3036	Cross-sectional	Public charities	USA (nationwide)	County	Organizational density	Age, employment, ethnicity, income	50
15	Kim and Kim (2015)	1971–2010	39	Time series	Registered nonprofits	South Korea (nationwide)	Country	Financial size	Gender, household location, income, origin	18
16	Kim and Kim (2016a)	1990–2012	463	Longitudinal	Nonprofits dominated by charitable donations	Cross-nation (OECD countries)	Country	Financial size	Gender, household location	4

Table 1 continued

#	Study	Year coverage	Sample size	Data structure	Type of nonprofits	Nonprofit country	Level of analysis	Measurement of nonprofit size	Measurement of population heterogeneity	# of effect size
17	Kim and Kim (2016b)	1992–2012	275	Longitudinal	Nonprofits dominated by charitable donations	Cross-nation (OECD countries)	Country	Financial size	Age, gender, household location, origin	20
18	Lee and Weisbrod (1977)	1960	48	Cross-sectional	Health care	USA (nationwide)	State	Financial size	Age, education, ethnicity, household location, income, religion	6
19	Liu (2016)	2002	3032	Cross-sectional	Public charities	USA (nationwide)	County	Organizational density, financial size	Education, employment, ethnicity, household location, income, religion	48
20	Marchesini da Costa (2016)	2001–2010	5562	Cross-sectional	Registered nonprofits	Brazil (nationwide)	Municipality	Organizational density	Employment, household location, income	15
21	Marcuello (1998)	1991	40	Cross-sectional	Registered nonprofits	Spain (Catalonia)	County	Organizational density	Age, education, employment	10
22	Matsunaga and Yamauchi (2004)	1992–1999	400	Longitudinal	Public charities	USA (nationwide)	State	Organizational density	Age, employment, ethnicity, household location	8
23	Matsunaga et al. (2010)	1995	44	Cross-sectional	Education and health	Cross-nation (CNP countries) ^a	Country	Employment size	Religion	4
24	Okten and Osili (2004)	1997–1998	3050	Cross-sectional	Community organizations	Indonesia (nationwide)	Community	Existence of organizations	Ethnicity, income	6
25	Onder (2011)	2000	35	Cross-sectional	Voluntary organizations	Cross-nation (35 countries)	Country	Voluntary participation	Ethnicity, income	2
26	Paarlberg and Gen (2009)	2003	1489	Cross-sectional	Public education	USA (San Francisco Bay)	School	Nonprofit coproduction	Ethnicity	4
27	Peck (2008)	1990–2000	656	Cross-sectional	Anti-poverty	USA (Phoenix, AZ)	Census tract	Organizational density, financial size	Education, ethnicity, employment	16
28	Pevcin (2012)	Not specified	38	Cross-sectional	Nonprofit sector	Cross-nation (CNP countries)	Country	Employment size, financial size	Income, language, ethnicity, religion	16
29	Polson (2017)	2010	3109	Cross-sectional	Anti-poverty	USA (nationwide)	County	Organizational density	Education, ethnicity, income, religion	8
30	Salamon and Anheier (1998)	Not specified	8	Cross-sectional	Nonprofit sector	Cross-nation (CNP countries)	Country	Employment size	Language	2
31	Salamon et al. (2000)	Not specified	22	Cross-sectional	Nonprofit sector	Cross-nation (CNP countries)	Country	Employment size	Religion	1

Table 1 continued

#	Study	Year coverage	Sample size	Data structure	Type of nonprofits	Nonprofit country	Level of analysis	Measurement of nonprofit size	Measurement of population heterogeneity	# of effect size
32	Smith (2013)	1965	351	Cross-sectional	Association	USA (Massachusetts)	Municipality	Organizational density	Ethnicity	1
33	Smith and Shen (2002)	1970–1994	124	Cross-sectional	Association	Cross-nation (124 countries)	County	Organizational density	Ethnicity, religion	2
34	Stater (2010)	1991–2003	33,016	Longitudinal	Public charities	USA (nationwide)	County	Financial size	Income	1
35	Stretesky et al. (2011)	1990–2008	3144	Cross-sectional	Environmental justices	USA (nationwide)	County	Organization founding	Education, ethnicity	12
36	Van Puyvelde and Brown (2016)	2012	254	Cross-sectional	Public charities	USA (Texas)	County	Organizational density	Age, education, employment, ethnicity, income	24
37	Wolch and Geiger (1983)	1978–1979	80	Cross-sectional	Social service	USA (Los Angeles county)	City	Organizational density	Age, ethnicity	12

^aCNP means the John Hopkins Comparative Nonprofit Sector Project (see <http://ccss.jhu.edu/research-projects/comparative-nonprofit-sector-project/> for more details about the dataset)

^bHousehold location means the ruralness or urbanness of a household's geographic location

bound estimates. (4) When original studies report multiple effect sizes, mostly because of different variable measurements, model specifications, or sample restrictions (e.g., Kim 2015; Liu 2016; Van Puyvelde and Brown 2016), all relevant effect sizes were coded to maintain within-study variation. Finally, 491 effect sizes were coded from 37 original studies. These effect sizes range from – 0.6815 in Kim and Kim (2015) to .9450 in James (1987). Within this group of effect sizes, there are 265 positive effects, 32 null effects, and 194 negative effects.

In addition to effect size information, we coded study-level descriptors such as characteristics of nonprofits under study and research design. Table 1 provides an overall picture of the distribution of the study descriptors across all 37 studies, which gives us a sense about the studies included in the meta-analysis.

Results and Findings

Table 2 reports the average effect sizes between population heterogeneity and nonprofit sector size under various scenarios. Before combining individual effect sizes across studies, we need to evaluate effect size heterogeneity to choose between a fixed-effects model and a random-effects model in the estimation of average effect size (Geyskens et al. 2009; Ringquist 2013). First, Hedge's *Q* test was performed to test the null hypothesis that the variation among the effect sizes could be accounted for by sampling error alone. Second, we calculated the *I*² statistic to further gauge the proportion of the variability in effect sizes that cannot be attributed to sampling error. After choosing a fixed-effects or a random-effects model, we estimated average effect sizes under different scenarios.⁷ In the estimation of average effect sizes, three methodological treatments are worth noting. First, meta-analysis does not treat the effect sizes from each original study equally, but it aggregates them with weights that are the inverse of the effect size variance (Ringquist 2013; Shadish and Haddock 2009). Effect sizes from larger-sample studies are thus weighted more heavily because such studies tend to produce estimators that are closer to the population. In this way, we accounted for the potential influence of sample size in the precision of average effect sizes across studies. Second, given that Fisher's *z* is not readily interpretable, average effect size was calculated in Fisher's *z* and then converted back to Pearson's *r* for reporting and interpretation (Borenstein 2009; Ringquist 2013). Third, we also performed *z* tests (two-tailed) to check whether average effect sizes are significantly different from zero, in order to

⁷ Stata command *metan* was used to test effect size heterogeneity and calculate weighted average effect sizes.

Table 2 Average effect sizes under different scenarios

Scenario	<i>k</i>	Effect size heterogeneity			Weighted average effect size		
		<i>Q</i> (<i>df</i>)	<i>I</i> ² (%)	Analytical model	<i>r</i>	<i>z</i>	95% CI
Overall	491	16,528.51*** (490)	97.0	Random-effects	.034	4.10***	[.018, .050]
<i>Country of nonprofit sector</i>							
U.S. nonprofits	369	15,530.08*** (368)	97.6	Random-effects	.028	2.86**	[.009, .048]
Non-U.S. nonprofits	122	884.82*** (121)	86.3	Random-effects	.043	3.84***	[.021, .065]
<i>Within-/cross-country comparison</i>							
Within-country	428	16,248.43*** (427)	97.4	Random-effects	.038	3.17**	[.011, .065]
Cross-country	63	276.56*** (62)	77.6	Random-effects	.055	3.53***	[.037, .073]
<i>Measurement of nonprofit sector size</i>							
Density	274	4517.97*** (273)	94.0	Random-effects	.021	2.98**	[.007, .035]
Finance	131	9794.34*** (130)	98.7	Random-effects	.022	3.36**	[.003, .041]
Employment	32	86.16*** (31)	64.0	Random-effects	.076	2.13*	[.006, .146]
Others	54	468.53*** (53)	88.7	Random-effects	.077	5.45***	[.049, .105]
<i>Measurement of population heterogeneity</i>							
Age	57	659.88*** (56)	91.5	Random-effects	.059	3.33***	[.024, .094]
Education	45	1224.42*** (44)	96.4	Random-effects	.072	2.77**	[.021, .123]
Employment status	65	485.72*** (64)	86.8	Random-effects	− .037	3.63***	[− .057, − .017]
Ethnicity	144	922.78*** (143)	84.6	Random-effects	.020	2.67**	[.005, .034]
Gender	13	24.98** (12)	52.0	Random-effects	− .120	3.56***	[− .186, .054]
Household location	34	608.10*** (33)	94.6	Random-effects	.033	1.52	[− .009, .075]
Income	65	658.98*** (64)	90.3	Random-effects	.016	1.44	[− .006, .038]
Language	11	6.94 (10)	0	Fixed-effects	.147	4.87***	[.118, .176]
Origin	15	11.65 (14)	0	Fixed-effects	.021	1.09	[− .017, .060]
Religion	42	674.38*** (41)	94.1	Random-effects	.147	5.26***	[.093, .202]

k number of effect sizes, *Q* Hedge's test of effect size heterogeneity, *df* degree of freedom, *I*² measure of effect size variability, *r* Pearson's *r*, *z* test of significance from zero, *CI* confidence interval

* significant at .05; ** significant at .01; *** significant at .001

assess the significance of the effects. The results are discussed as follows.

We first estimated the average effect size for all 491 effect sizes coded from the 37 existing studies. The Q test was rejected at the .001 level, implying significant variability in effect sizes that is not attributable to sampling error. Again, an I^2 statistic of 97% indicates a high level of heterogeneity across effect sizes (Higgins and Thompson 2002). Based on these findings, a random-effects model is more appropriate to synthesize the effect sizes. The weighted average effect size is .034 [95% CI (.18, .50), $z = 4.10$, $p < .001$]. This effect size indicates a significant and positive association between population heterogeneity and nonprofit sector size, but the magnitude of the relationship is substantially small.⁸ This finding implies two things regarding the demand heterogeneity hypothesis. First, the positive sign of the average effect size supports the hypothesis that higher degrees of population heterogeneity in a locality could be associated with more nonprofit formation and growth. Second, the very small magnitude of the average effect size questions the robustness of the hypothesis, as it shows that population heterogeneity only has a trivial relationship with nonprofit sector size and there must be other factors shaping nonprofit formation and growth in a more forceful manner. Put together, although the demand heterogeneity hypothesis still holds, its explanatory power in predicting nonprofit sector size seems less strong.

Table 2 also reports average effect sizes under other scenarios to check the moderating effects of different research design components on the test of the demand heterogeneity hypothesis. Similarly, we began with a choice between a fixed-effects and a random-effects model. In most scenarios, random-effects models were used to estimate average effect sizes because there are significant variabilities among effect sizes. Two exceptions come from effect sizes related to linguistic and origin heterogeneity where limited effect size variations were found. In these two cases, fixed-effect models were employed to aggregate effect sizes. Then, weighted average effect size in each scenario was calculated and discussed as follows.

First, it could be argued that each country has its unique institutional and cultural environment and thus makes itself less comparable to others (Abzug and Turnheim 1998; Salamon and Anheier 1998). In this line of reasoning, whether the U.S.-based demand heterogeneity hypothesis could be applied to non-U.S. countries is also an open question. Actually, existing tests of the demand

heterogeneity hypothesis are dominated by studies on U.S. nonprofits. In our meta-analysis, 369 effect sizes come from studies on U.S. nonprofits, and 122 effect sizes are from studies on non-U.S. nonprofits. We estimated average effect sizes for these two groups of studies separately.⁹ In studies specifically looking at U.S. nonprofits, the weighted average effect size is .028 [95% CI (.009, .048), $z = 2.86$, $p < .01$], while in studies including non-U.S. nonprofits, their weighted average effect size is .043 [95% CI (.021, .065), $z = 3.84$, $p < .001$]. Both average effect sizes point to a significant and positive association between population heterogeneity and nonprofit sector size, even though such an effect is relatively small. In this way, the demand heterogeneity hypothesis seems to be applicable to not only U.S. but also non-U.S. nonprofits. As such, the hypothesis can be supported with less reference to country-specific environments.

Similarly, it could also be argued that because each country might be less comparable, studies employing cross-country data might compare apples and oranges and thus produce different results. In our sample of 491 effect sizes, 428 effect sizes are from studies focusing on one specific country (within-country studies), while 63 effect sizes are from studies relying on cross-country data such as OECD data and United Nation data (cross-country studies). We compared the average effect sizes from these two groups of studies. In within-country studies, the weighted average effect size is .038 [95% CI (.011, .065), $z = 3.17$, $p < .01$], and in cross-country studies, their weighted average effect size is .055 [95% CI (.037, .073), $z = 3.53$, $p < .001$]. There seems no substantial difference between these two effect sizes: both support the demand heterogeneity hypothesis and their magnitudes are rather small. This finding implies that the demand heterogeneity hypothesis can be supported by both within-country and cross-country data.

Third, we explored whether the use of different measurements of nonprofit sector size would influence the test of the demand heterogeneity hypothesis. Among the 37 studies included in the meta-analysis, 3 predominant measurements of nonprofit sector size are used, namely organizational density (number of nonprofits within a geographic area), financial size (nonprofit revenue, asset, or

⁸ According to the rule of thumb in interpreting correlation coefficients, a correlation coefficient r between 0 and .2 is considered as a very weak positive association between two variables (e.g., Cohen 1988).

⁹ Because of limited studies on non-U.S. nonprofits, we could not estimate an average effect size for each non-U.S. country, but only calculate an average effect size for non-U.S. countries as a whole. For example, in our sample of 491 effect sizes, there are only one study contributing 6 effect sizes on Japanese nonprofits (Kanaya et al. 2015), one study producing 6 effect sizes on Indonesian nonprofits (Okten and Osili 2004), and one study with 10 effect sizes on Spanish nonprofits (Marcuello 1998). Aggregating a very small number of effect sizes from a single study might not be able to reach a meaningful average effect size, as meta-analysis is generally used to integrate multiple effect sizes from different studies.

expenditure), and employment size (full-time equivalent nonprofit employees). We calculated an average effect size for each group of effect sizes with one specific measurement. In effect sizes with organizational density measurement, the weighted average effect size is .021 [95% CI (.007, .035), $z = 2.98$, $p < .01$], in effect sizes employing financial size measurement, their weighted average effect size is .022 [95% CI (.003, .041), $z = 3.36$, $p < .01$], in effect sizes relying on employment size measurement, their weighted average effect size is .076 [95% CI (.006, .146), $z = 2.13$, $p < .05$], and in effect sizes representing other measurements, their weighted average effect size is .077 [95% CI (.049, .105), $z = 5.45$, $p < .001$]. Considering this evidence, how nonprofit sector size is measured seems not to challenge the demand heterogeneity hypothesis, for each of the average effect sizes, though small, is significant and positive. In short, the measurement of nonprofit sector size makes no difference in the support of the hypothesis.

Fourth, we examined the validity of each dimension of population heterogeneity in predicting nonprofit sector size. Although Weisbrod suggested that heterogeneity in terms of age, income, education, religion, ethnicity, or other demographic variables can be used as indicators of population heterogeneity to represent demand heterogeneity, he did admit that “we do not have any strong a priori basis for selecting these particular indices of demand heterogeneity” (Lee and Weisbrod 1977, p. 85). Indeed, existing empirical evidence suggests that not every type of population heterogeneity could predict nonprofit growth equally well (see Kim 2015; Matsunaga et al. 2010 for brief reviews of this issue). We calculated average effect sizes for each group of effect sizes representing one specific dimension of population heterogeneity. The finding is twofold. First, there are five indicators of population heterogeneity (age, education, ethnicity, language, and religion) that support the demand heterogeneity hypothesis well, as the average effect size in each group is significant and positive. Second, another five indicators (employment status, gender, household location, income, and origin) fail to support the demand heterogeneity hypothesis because their average effect sizes are either insignificant or negative, contrary to what the hypothesis predicts.¹⁰ In this way, the demand heterogeneity hypothesis could be supported when population heterogeneity is measured in terms of age, education, ethnicity, language, and religion.

¹⁰ It should be noted that although these five effect sizes seem to be different from what the demand heterogeneity hypothesis predicts, they do not necessarily mean the demand heterogeneity hypothesis should be rejected. Rather, they only imply that the demand heterogeneity hypothesis is not supported when population heterogeneity is measured in terms of employment status, gender, household location, income, and origin, but supported by other measures of population heterogeneity.

In sum, all the weighted average effect sizes underlying the demand heterogeneity hypothesis reported in Table 2 tend to imply: (1) the relationship between population heterogeneity and nonprofit sector size is overall significant and positive; (2) the magnitude of the effect is generally very small; (3) the hypothesis has been supported by both U.S. and non-U.S. data, as well as within-country and cross-country data; (4) the hypothesis can be supported when nonprofit sector size is measured by organizational density, financial size, or employment size; and (5) the hypothesis could be supported if population heterogeneity is measured in terms of age, education, ethnicity, language, or religion. To conclude, the demand heterogeneity hypothesis is supported by existing empirical evidence, but its explanatory power might be relatively weak.

Discussion and Conclusion

The nonprofit sector plays a vital and unique role within the good governance of a society by addressing various social, economic, and political needs that neither business nor government is prepared to fulfill. Over the past several decades, we have witnessed a substantial growth of the nonprofit sector in various countries around the world, in both size and impact. However, despite the unprecedented growth, the size of the nonprofit sector and its growth patterns vary dramatically across localities. What accounts for the size variation of the nonprofit sector, or to frame it differently, why nonprofits exist, has attracted much scholarly attention over the years. Plenty of theories have been developed to address this issue. One dominant theory among this body of literature is the government failure theory proposed by Weisbrod. A key model of the theory suggests that nonprofits exist to fill the service gap left by government provision, largely because of the conflict between diversified needs due to population heterogeneity and government underprovision due to majority voting. As such, there should be a positive relationship between one locality's degree of population heterogeneity and its nonprofit sector size. This demand heterogeneity hypothesis has received wide recognition as a classic work in nonprofit studies and been tested extensively over decades.

However, recent studies of the size variation of the nonprofit sector reach an inconsistent conclusion concerning the relationship between population heterogeneity and nonprofit sector size. Although some studies do find the expected positive association between the two, many other studies in this line of research show either null or negative association and consequently question the explanatory power of Weisbrod's demand heterogeneity hypothesis. Indeed, these competing findings fail to provide a coherent assessment of the hypothesis and thus cause a significant

barrier for future nonprofit studies. We argue in the present study that a pressing scholarly need is to take stock of these existing studies and draw a generalized conclusion from their competing findings.

Toward this goal, we conducted a meta-analysis to aggregate existing literature testing the demand heterogeneity hypothesis. Through a systematic review of 37 existing empirical studies with 491 effect sizes, we find an average correlation between population heterogeneity and nonprofit sector size ranging from .020 to .147. This finding of average effect size has two theoretical implications. First, we could confirm the existence of a positive association between a locality's degree of population heterogeneity and its nonprofit sector size. Therefore, the size of the nonprofit sector will be larger where the degree of population heterogeneity is higher. In this way, the present study provides meta-analytical support for Weisbrod's demand heterogeneity hypothesis. Indeed, although the hypothesis has been criticized for ignoring the supply-side effect of population heterogeneity and there is empirical evidence showing that population heterogeneity can be detrimental to nonprofit formation, the positive net effect we found here implies that the positive demand-side effect of population heterogeneity on nonprofit formation seems to outweigh its negative supply-side effect. Overall, population heterogeneity would still push nonprofits to take root and grow.

Second, the magnitude of the association indicates that population heterogeneity might only have a limited relationship with nonprofit sector size, which further questions the robustness of the hypothesis. Admittedly, meta-analysis cannot delve into causal factors, and thus we are not able to determine the causality between population heterogeneity and nonprofit sector size in this study. However, the weak association between the two at least implies that population heterogeneity might not be a strong factor shaping nonprofit sector size; there must be other more important factors causing nonprofits' establishment and growth. In this way, although the meta-analytical result generally supports Weisbrod's line of reasoning, the power of this hypothesis in explaining and predicting nonprofit sector size might be less robust. Future studies need to be cautious when applying the government failure theory to explore the size variation of the nonprofit sector by locality and give more attention to other theories on nonprofit formation and behaviors.

Further, we also explored a number of methodological factors to examine whether these factors would influence empirical supports of the demand heterogeneity hypothesis. First, the hypothesis seems to be applicable to both U.S. and non-U.S. nonprofits and thus can be generalized across countries. Second, the hypothesis can be supported by both within- and cross-country data, and thus data type seems

not to make a difference. Third, different measurements of nonprofit sector size do not undermine the validity of the hypothesis. As such, organizational density, financial capacity, or employment size can be used to gauge nonprofit sector size. Fourth, some indicators of population heterogeneity (age, education, ethnicity, language, and religion) seem to support the hypothesis better than other indicators (employment status, gender, household location, income, and origin). Future studies might rely more on the first five proxies. In sum, these findings should be able to inform future application of the demand heterogeneity hypothesis in the exploration of nonprofit sector size variation.

In addition to these contributions, the current research is subject to some limitations which could be addressed by future studies. On the theoretical side, first, existing studies are dominated by studies on U.S. nonprofits. Although the present analysis finds the average effect size from U.S. studies to be roughly the same as that from non-U.S. studies entirely, we cannot guarantee the demand heterogeneity hypothesis' applicability to every non-U.S. country.¹¹ In this sense, more studies on other non-U.S. countries and cross-national comparison could contribute to our understanding of the topic. Before that, generalization to other non-U.S. countries should still proceed on a case-by-case basis considering a country's specific environment. Second, the present analysis explores the association between population heterogeneity and nonprofit sector size, without including government provision as an intervening variable. A more thorough test of Weisbrod's model might take the mediating effect of government social spending into consideration.¹² Third, existing studies generally use one demographic dimension (such as age, race, education) to measure population heterogeneity as a multidimensional construct. Although we find that some dimensions support the demand heterogeneity hypothesis better than others, future studies might consider developing a more comprehensive index to capture the multidimensional nature of population heterogeneity (e.g., Lu 2016). Further, population heterogeneity (observed) is usually resorted to as a proxy of demand population heterogeneity

¹¹ The Weisbrod's theory is deeply rooted in majoritarian democracy settings where minority interests are poorly represented in the political system and thus nonprofits are needed to meet unmet demand for minority-preferred public goods. However, in many other countries with a consensus democracy system, Weisbrod's causal model might not apply. The consensus democracy system provides proportional political representation to minority interest and a failure to deliver public goods for these minorities would reduce politicians' chances of being elected. As a consequence, population heterogeneity could cause public spending to increase to satisfy diverse minority preferences, which may leave less room for nonprofits to fill. We thank one anonymous reviewer for this observation.

¹² We are obliged to one anonymous reviewer for this comment.

(unobserved), but the validity of this measurement is still not for sure.

On the methodological side, it should be first noted that any meta-analysis is limited by the availability of empirical studies. Therefore, the current meta-analysis only adjusted for the effect of sample size in the aggregation of average effect sizes using a traditional meta-analysis method (Ringquist 2013; Shadish and Haddock 2009). Although the effects of other statistical and methodological artifacts such as measurement reliability and variable ranges have been identified (Schmidt and Hunter 2015), existing studies generally do not provide sufficient information in these regards, which makes it impossible for us to make additional adjustments. Second, existing studies heavily rely on cross-sectional data, and thus the limitations related to cross-sectional analysis would apply to the meta-analysis here.¹³ Clearly, without longitudinal analysis of the relationship between population heterogeneity and nonprofit sector size, we fail to account for the interaction between the two variables over time. Third, many existing studies focus on the entire nonprofit sector or aggregate several subsectors together without accounting for the subsector variation. Some studies do demonstrate the existence of subsector variation due to different policy and resource environments (e.g., Bielefeld et al. 1997; Grønbyrg and Paarlberg 2001). In this way, within-subsector analysis and cross-subsector comparison should attract more scholarly attention.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abzug, R., & Turnheim, J. K. (1998). Bandwagon or band-aid? A model of nonprofit incorporation by state. *Nonprofit and Voluntary Sector Quarterly*, 27(3), 300–322.
- Alesina, A., & La Ferrara, E. (2000). Participation in heterogeneous communities. *Quarterly Journal of Economics*, 115(3), 847–904.
- Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, 85(2), 207–234.
- Anderson, C. J., & Paskeviciute, A. (2006). How ethnic and linguistic heterogeneity influence the prospects for civil society: A comparative study of citizenship behavior. *Journal of Politics*, 68(4), 783–802.
- Andreoni, J., Payne, A. A., Smith, J., & Karp, D. (2016). Diversity and donations: The effect of religious and ethnic diversity on charitable giving. *Journal of Economic Behavior & Organization*, 128, 47–58.
- Bendick, M. (1977). Education as a three-sector industry. In B. A. Weisbrod (Ed.), *The voluntary nonprofit sector: An economic analysis* (pp. 101–142). Lexington, MA: D.C. Heath and Compan.
- Ben-Ner, A., & Van Hoomissen, T. (1991). Nonprofit organizations in the mixed economy. *Annals of Public and Cooperative Economics*, 62(4), 519–550.
- Bielefeld, W., & Murdoch, J. C. (2004). The locations of nonprofit organizations and their for-profit counterparts: An exploratory analysis. *Nonprofit and Voluntary Sector Quarterly*, 33(2), 221–246.
- Bielefeld, W., Murdoch, J. C., & Waddell, P. (1997). The influence of demographics and distance on nonprofit location. *Nonprofit and Voluntary Sector Quarterly*, 26(2), 207–225.
- Buchanan, J. M., & Tullock, G. (1962). *The calculus of consent*. Ann Arbor, MI: University of Michigan Press.
- Burger, A., & Veldheer, V. (2001). The growth of the nonprofit sector in the Netherlands. *Nonprofit and Voluntary Sector Quarterly*, 30(2), 221–246.
- Cohen, J. (1988). *Statistical power analysis for the behavior science*. Hillsdale, NJ: Lawrence Erlbaum Association.
- Corbin, J. J. (1999). A study of factors influencing the growth of nonprofits in social services. *Nonprofit and Voluntary Sector Quarterly*, 28(3), 296–314.
- Costa, D. L., & Kahn, M. E. (2003). Civic engagement and community heterogeneity: An economist's perspective. *Perspectives on Politics*, 1(1), 103–111.
- DiMaggio, P. J., & Anheier, H. K. (1990). The sociology of nonprofit organizations and sectors. *Annual Review of Sociology*, 16, 137–159.
- Douglas, J. (1987). Political theories of nonprofit organization. In W. W. Powell (Ed.), *The nonprofit sector: A research handbook* (pp. 43–54). New Haven, CT: Yale University Press.
- Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237–254). New York: Russell Sage Foundation.
- Fong, C. M., & Luttmmer, E. F. P. (2009). What determines giving to Hurricane Katrina victims? Experimental evidence on racial group loyalty. *American Economic Journal: Applied Economics*, 1(2), 64–87.
- Garrow, E. E., & Garrow, S. H. (2014). White flight and the presence of neighborhood nonprofit organizations: Ethno-racial transition, poverty, and organizational resources. *Race and Social Problems*, 6(4), 328–341.
- Geyskens, I., Krishnan, R., Steenkamp, J. B. E., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management*, 35(2), 393–419.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, 115(3), 811–846.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8.
- Grønbyrg, K. A., & Paarlberg, L. (2001). Community variations in the size and scope of the nonprofit sector: Theory and preliminary findings. *Nonprofit and Voluntary Sector Quarterly*, 30(4), 684–706.
- Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review*, 101(04), 709–725.
- Hansmann, H. B. (1980). The role of nonprofit enterprise. *Yale Law Journal*, 89(5), 835–901.
- Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558.

¹³ Within the 37 studies included in the meta-analysis, only four studies employ panel data analysis (Kanaya et al 2015; Kim and Kim 2016a, b, Matsunaga and Yamauchi 2004) and one study uses time-series data analysis (Kim and Kim 2015).

- James, E. (1987). The public/private division of responsibility for education: An international comparison. *Economics of Education Review*, 6(1), 1–14.
- James, E. (1993). Why do different countries choose a different public-private mix of educational services? *Journal of Human Resources*, 28(3), 571–592.
- Joassart-Marcelli, P. (2013). Ethnic concentration and nonprofit organizations: The political and urban geography of immigrant services in Boston, Massachusetts. *International Migration Review*, 47(3), 730–772.
- Joassart-Marcelli, P., & Wolch, J. R. (2003). The intrametropolitan geography of poverty and the nonprofit sector in Southern California. *Nonprofit and Voluntary Sector Quarterly*, 32(1), 70–96.
- Kanaya, N., Takahashi, H., & Shen, J. (2015). The market share of nonprofit and for-profit organizations in the quasi-market: Japan's long-term care services market. *Annals of Public and Cooperative Economics*, 86(2), 245–266.
- Kim, M. (2015). Socioeconomic diversity, political engagement, and the density of nonprofit organizations in US counties. *American Review of Public Administration*, 45(4), 402–416.
- Kim, S. E., & Kim, Y. H. (2015). Measuring the growth of the nonprofit sector: A longitudinal analysis. *Public Administration Review*, 75(2), 242–251.
- Kim, Y. H., & Kim, S. E. (2016a). Testing an economic model of nonprofit growth: Analyzing the behaviors and decisions of nonprofit organizations, private donors, and governments. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 27(6), 2937–2961.
- Kim, Y. H., & Kim, S. E. (2016b). What accounts for the variations in nonprofit growth? A cross-national panel study. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*. <https://doi.org/10.1007/s11266-016-9752-x>.
- Kingma, B. R. (2003). Public good theories of the nonprofit sector. In H. K. Anheier & A. Ben-Ner (Eds.), *The study of the nonprofit enterprise* (pp. 53–65). New York: Springer.
- Lecy, J. D., & Van Slyke, D. M. (2013). Nonprofit sector growth and density: Testing theories of government support. *Journal of Public Administration Research and Theory*, 23(1), 189–214.
- Lee, A. J., & Weisbrod, B. A. (1977). Collective goods and the voluntary sector: The case of the hospital industry. In B. A. Weisbrod (Ed.), *The voluntary nonprofit sector: An economic analysis* (pp. 77–100). Lexington, MA: D.C. Heath and Compan.
- Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 147–158). New York: Russell Sage Foundation.
- Liu, G. (2016). Government decentralization and the size of the nonprofit sector: Revisiting the government failure theory. *American Review of Public Administration*. <https://doi.org/10.1177/02750740166631643>.
- Lu, J. (2016). Governing diversity: The impact of social heterogeneity on human service delivery. *Public Organization Review*, 16(2), 153–166.
- Lyons, M. (1993). The history of non-profit organisations in Australia as a test of some recent non-profit theory. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 4(3), 301–325.
- Marchesini da Costa, M. (2016). What influences the location of nonprofit organizations? A spatial analysis in Brazil. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 27(3), 1064–1090.
- Marcuello, C. (1998). Determinants of the non-profit sector size: An empirical analysis in Spain. *Annals of Public and Cooperative Economics*, 69(2), 175–192.
- Matsunaga, Y., & Yamauchi, N. (2004). Is the government failure theory still relevant? A panel analysis using US state level data. *Annals of Public and Cooperative Economics*, 75(2), 227–263.
- Matsunaga, Y., Yamauchi, N., & Okuyama, N. (2010). What determines the size of the nonprofit sector? A cross-country analysis of the government failure theory. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 21(2), 180–201.
- McKeever, B. S. (2015). *The nonprofit sector in brief 2015*. Washington, DC: The Urban Institute.
- Miguel, E., & Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of public Economics*, 89(11), 2325–2368.
- Okten, C., & Osili, U. O. (2004). Contributions in heterogeneous communities: Evidence from Indonesia. *Journal of Population Economics*, 17(4), 603–626.
- Onder, M. (2011). A preliminary cross-national test of competing theories of nonprofits: Does culture matter? *International Review of Public Administration*, 16(1), 71–90.
- Paarlberg, L. E., & Gen, S. (2009). Exploring the determinants of nonprofit coproduction of public service delivery: The case of K-12 public education. *American Review of Public Administration*, 39(4), 391–408.
- Peck, L. R. (2008). Do antipoverty nonprofits locate where people need them? Evidence from a spatial analysis of Phoenix. *Nonprofit and Voluntary Sector Quarterly*, 37(1), 138–151.
- Pevcin, P. (2012). Analysis of cross-country differences in the nonprofit sector size. *Prague Economic Papers*, 2, 186–204.
- Polson, E. C. (2017). Religious environments and the distribution of anti-poverty nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, 46(1), 156–174.
- Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century. *Scandinavian Political Studies*, 30(2), 137–174.
- Reed, J. G., & Baxter, P. M. (2009). Using reference databases. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 73–101). New York: Russell Sage Foundation.
- Ringquist, E. (2013). *Meta-analysis for public management and policy*. San Francisco: Wiley.
- Rose-Ackerman, S. (1996). Altruism, nonprofits, and economic theory. *Journal of Economic Literature*, 34(2), 701–728.
- Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 103–126). New York: Russell Sage Foundation.
- Salamon, L. M. (1987). Of market failure, voluntary failure, and third-party government: Toward a theory of government-nonprofit relations in the modern welfare state. *Nonprofit and Voluntary Sector Quarterly*, 16(1–2), 29–49.
- Salamon, L. M., & Anheier, H. K. (1998). Social origins of civil society: Explaining the nonprofit sector cross-nationally. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 9(3), 213–248.
- Salamon, L. M., Sokolowski, S. W., & Anheier, H. K. (2000). *Social origins of civil society: An overview*. Baltimore, MD: Johns Hopkins University Center for Civil Society Studies.
- Salamon, L. M., Sokolowski, S. W., Haddock, M. A., & Tice, H. S. (2013). *The state of global civil society and volunteering: Latest findings from the implementation of the UN nonprofit handbook*. Baltimore, MD: Johns Hopkins University Center for Civil Society Studies. Working paper #49.
- Saxton, G. D., & Benson, M. A. (2005). Social capital and the growth of the nonprofit sector. *Social Science Quarterly*, 86(1), 16–35.

- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks: Sage.
- Shadish, W., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 257–278). New York: Russell Sage Foundation.
- Smith, D. H. (2013). Explaining the prevalence of voluntary associations among 351 Massachusetts municipalities circa 1965: Testing a theory. *The China Nonprofit Review*, 5(1), 177–193.
- Smith, D. H., & Shen, C. (2002). The roots of civil society: A model of voluntary association prevalence applied to data on larger contemporary nations. *International Journal of Comparative Sociology*, 43(2), 93–133.
- Stater, K. J. (2010). How permeable is the nonprofit sector? Linking resources, demand, and government provision to the distribution of organizations across nonprofit mission-based fields. *Nonprofit and Voluntary Sector Quarterly*, 39(4), 674–695.
- Steinberg, R. (2006). Economic theories of nonprofit organizations. In W. W. Powell & R. Steinberg (Eds.), *The nonprofit sector: A research handbook* (2nd ed., pp. 117–139). New Haven, CT: Yale University Press.
- Stretesky, P. B., Huss, S., Lynch, M. J., Zahran, S., & Childs, B. (2011). The founding of environmental justice organizations across US counties during the 1990s and 2000s: Civil rights and environmental cross-movement effects. *Social Problems*, 58(3), 330–360.
- Van Puyvelde, S., & Brown, W. A. (2016). Determinants of nonprofit sector density: A stakeholder approach. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 27(3), 1045–1063.
- Weisbrod, B. A. (1975). Toward a theory of the voluntary non-profit sector in a three sector economy. In E. S. Phelps (Ed.), *Altruism, morality and economic theory* (pp. 171–296). New York: Russell Sage Foundation.
- Weisbrod, B. A. (1977a). Toward a theory of the voluntary non-profit sector in a three sector economy. In B. A. Weisbrod (Ed.), *The voluntary nonprofit sector: An economic analysis* (pp. 51–76). Lexington, MA: D.C. Heath and Compan.
- Weisbrod, B. A. (Ed.). (1977b). *The voluntary nonprofit sector: An economic analysis*. Lexington, MA: D.C. Heath and Compan.
- Weisbrod, B. A. (1986). Toward a theory of the voluntary non-profit sector in a three sector economy. In S. Rose-Ackerman (Ed.), *The economics of nonprofit institutions* (pp. 21–44). New York: Oxford University Press.
- Weisbrod, B. A. (1988). *The nonprofit economy*. Cambridge, MA: Harvard University Press.
- Wilson, J. Q. (2000). *Bureaucracy: What government agencies do and why they do it*. New York: Basic Books.
- Wolch, J. R., & Geiger, R. K. (1983). The distribution of urban voluntary resources: An exploratory analysis. *Environment and Planning A*, 15(8), 1067–1082.