CrossMark

Audit firms face downward-sloping demand curves and the audit market is far from perfectly competitive

Joseph Gerakos[1] (ID) · **Chad Syverson[2]**

Abstract We discuss the discrete choice demand estimation methods applied by Guo et al. (2017) in the audit setting. We then review insights into audit market competition that demand estimation has already provided. We conclude that the audit market is far from perfectly competitive.

Keywords Auditing · Demand estimation · Competititon

JEL Classification M42 · L84

1 Introduction

Guo et al. (2017) use *demand estimation* techniques to investigate how the introduction of a joint audit policy (requiring two audit firms to sign off on a client's audit) would affect the demand side of the audit market. Specifically, they estimate demand for the Big 4 and middle market audit firms under the French joint audit regime. They then apply these demand parameter estimates to the United Kingdom to estimate the change in consumer surplus under the counterfactual scenario in which the

✉ Joseph Gerakos
  joseph.j.gerakos@tuck.dartmouth.edu

[1] Tuck School of Business at Dartmouth College, 100 Tuck Hall, Hanover, NH 03755, USA

[2] University of Chicago Booth School of Business and NBER, Chicago, IL USA

United Kingdom implements a joint audit regime.[1] Their structural approach provides insights into this proposed policy change that would be missed in a reduced form analysis.

Their empirical demand estimation approach is similar to the methods used by Gerakos and Syverson (2015). In that study, we estimated audit firm demand in the United States to calculate changes in consumer surplus that would occur if mandatory audit firm rotation were implemented and if one of the Big 4 disappeared. Although demand estimation is new to the accounting literature, it—and in particular the discrete choice form applied in Guo et al. (2017) and Gerakos and Syverson (2015)—is common in many fields in economics (and industrial organization in particular).[2]

In this article, we explore how demand estimation can be applied in auditing research and review insights into the market that it has already provided. We set the foundation for the discussion by analyzing difficulties in the interpretation of the audit fee regression, a commonly applied tool in the auditing literature. We then discuss the mechanics of the discrete choice demand estimation approach used by Gerakos and Syverson (2015) and Guo et al. (2017). We go on to explain how the findings from this work imply that audit firms have market power. That is, the market is not perfectly competitive and, indeed, is far from it.

2 Interpretation of the audit fee regression

Instead of discrete choice demand estimation, the auditing literature has relied on the audit fee regression.[3] This approach typically involves researchers regressing the logarithm of audit fees on variables that capture client characteristics (e.g., size, profitability, industrial segments, foreign sales, litigation risk. . .). The estimated coefficients from these hedonic price regressions are then typically interpreted as capturing "supply-side" effects—the marginal costs of providing the audit. For example, studies typically interpret a positive and significant coefficient on a proxy for litigation risk as showing that it is more costly for audit firms to provide attestation services to clients that are subject to higher likelihoods of litigation.

This approach, its widespread use notwithstanding, is problematic. The reason is that the observed price in a market generally reflects both demand-side and supply-side influences on price. If a characteristic of a client firm could affect both audit demand and costs—and there are many examples of instances where this is likely—then the standard approach of regressing fees on those characteristics cannot

---

[1]In this setting, the change in consumer surplus can be interpreted as the aggregate dollar amount that audit clients would be willing to pay to not be subject to a joint audit regime. This measure does not capture changes in producer surplus or externalities imposed on entities outside the market. Thus it does not capture the full welfare effects of such a regulatory change.

[2]There are a wealth of resources explaining the basic concepts and practices of discrete choice demand estimation. For example, Anderson et al. (1992) and chapter 3 of Train (2009) provide thorough explanations of these methods.

[3]For a thorough review of this literature, see Hay et al. (2006).

reveal whether the relationship between the two reflects demand-side or supply-side influences.[4]

Simunic ([1980]) illustrates this issue in his study of whether the Big 8 audit firms exercise market power. He regresses audit fees on client characteristics along with an indicator for Big 8 firms. To interpret the coefficient on the Big 8 as a measure of market power, he develops a theoretical model that makes assumptions about both the supply- and demand-sides of the audit market. These assumptions allow him to interpret the coefficient on the Big 8 indicator variable as the extent that these firms exercise market power relative to the non-Big 8 audit firms.[5] This interpretation is valid within the narrow confines built by the assumptions of his model, but we demonstrate here that pitfalls exist in more general settings.

To make this issue transparent, let us continue with the example of the influence of litigation risk on observed audit fees. Suppose audit demand and costs exhibit the following very simple structure (unrealistically simple, but it is just to make the exposition clear).

We assume, as in the discussion above, that the marginal cost of an audit rises with litigation risk:

$$MC = C + bLR, \tag{1}$$

where MC is marginal cost, C a constant, and b parameterizes how quickly costs rise with litigation risk, LR.

Suppose now that *demand* for an audit also rises with litigation risk. This is plausible; firms with more potential for litigation are likely to be willing to pay more for a given level of audit quality if this helps reduce the probability of costly litigation. Specifically, we assume an audit firm faces the following demand curve for its audit services:

$$Q = A + dLR - eP, \tag{2}$$

where A is a constant, d parameterizes the relation between litigation risk and a client firm's demand for audit services, P is the audit fee, and e embodies the sensitivity of audit demand to fees. In the audit market, each client firm has demand for only one audit. However, we can interpret the quantity demanded Q as the probability that the particular client firm chooses the audit firm facing that demand curve. All else equal, this probability decreases with the audit fee. Thus the demand curve is downward sloping, even though, in the end, the client only purchases (at most) one audit (in regions without joint audit requirements). The discrete choice demand system we describe below works exactly the same way. It too makes "quantity demanded" a probabilistic value.

For ease of exposition, we set the sensitivity of audit demand to fees, e, equal to one. The marginal revenue curve corresponding to this demand curve would then be

$$MR = A + dLR - 2Q. \tag{3}$$

---

[4]For a detailed discussion of the interpretation of hedonic price regressions like the standard audit fee regression, see Rosen ([1974]).

[5]Note that he assumes that the Big 8 market and non-Big 8 markets are segmented and that the non-Big 8 market is perfectly competitive.

A firm with pricing power will set fees to equate marginal revenue with its marginal cost. Doing this for the expressions above and solving for the *P* that maximizes the audit firm's expected profit gives the equilibrium audit fee:

$$P = \frac{A + C}{2} + \frac{b + d}{2} LR. \tag{4}$$

This expression makes clear that the effect of litigation risk on audit fees comes from *both* its influence on costs (through *b*) *and* its influence on audit demand (through *d*). Audit fees are higher for high litigation risk firms in part because it is more costly to audit them but also because that risk raises their demand for audit services.

Recognize that the standard literature practice of regressing fees on litigation risk would not allow one to separate these two influences. In the expression above, for instance, the coefficient on litigation risk would be $(b + d)/2$. Interpreting this value as reflecting only the cost-side relationship would be incorrect. Of course, interpreting this value as reflecting only the demand-side relationship would be equally mistaken.

This is an example of a more general difficulty in interpreting audit fee regressions. As the expression above shows, both cost and demand factors influence observed fees more generally, through *C* from the cost side and *A* from the demand side. Any other characteristic of the audit that might affect both audit costs and client firms' demand for audit services (such as whether the client's shares are cross-listed, or its location, size, complexity, governance structure, whether its shares are publicly listed, the incentives of its executives, and so on) will lead to a similar conundrum in interpreting the underlying reasons for variation in audit fees.[6]

This problem, and its solution, are closely related to the issue of obtaining an unbiased estimate of the price sensitivity of demand, which we discuss below. There, the potential for supply-side responses to demand shifts (such as an audit firm raising fees to potential clients it knows to have idiosyncratically strong demand for its services) could create price and quantity (choice probability) variation that confounds supply and demand responses. That is, the price-choice combinations observed in the data mix supply and demand together. The same problem exists here: observed fees commingle supply and demand.

To be able to separately measure supply and demand effects on audit fees, one needs to estimate audit demand and supply conditions separately. This would isolate *b* from *d* and *A* from *C*. Fortunately, this is an old and long-solved (at least conceptually) problem in the economics literature. Indeed, as discussed by Angrist and Krueger (2001), the development of two-stage least squares (i.e., instrumental variables) methods in the 1920s was specifically driven by the need to separately identify the demand and supply curves from observed equilibrium prices and quantities. These methods use an instrument—a factor that shifts the supply curve but not the demand curve—to isolate price and quantity variation that only reflects the shape of the demand curve. This allows one to trace out and estimate just the demand

---

[6]We are not the first to make this point in the auditing setting. See, for example, Gaver and Gaver (1995), Copley et al. (1995), and Deis and Hill (1998), and Hay et al. (2006).

curve. Once the demand curve is estimated, it can be used in turn to back out the supply curve from observed price-choice combinations. Then, with separate measures of both the demand- and supply-side relationships between audit characteristics and fees in hand, the researcher can correctly interpret the sources of fee variation. In this way, the audit fee regression is exactly the kind of specification that the two-stage least squares approach was developed to estimate.

3 Discrete choice demand estimation

Estimating demand is therefore a crucial step in interpreting how client attributes affect audit fees. Yet this is just an example. It can in fact be a basic building block in a much broader swath of analyses into the causes and consequences of outcomes in the audit market. In this section, we turn our attention toward a form of demand estimation that is both practical and empirically powerful: discrete choice analysis.

In contrast with the typical audit fee approach, Guo et al. (2017) estimate a utility function:

$$U_{ij} = \delta_j - \alpha \ln(p_{ij}) + \beta_{ij} x_{ij} + \epsilon_{ij}, \tag{5}$$

which consists of a fixed effect for audit firm j, δ_j, the price of an audit of client i by audit firm j, p_{ij}, observable characteristics of audit firm j and audit client i that vary across the client's potential audit firm choices, x_{ij}, and an error term ϵ_{ij}. This equation represents an indirect utility function, of course, because it includes the product's price.[7] Like any utility function, it relates consumer choices to the consumer's utility level. It does not just embody a single demand curve for a particular good but instead embodies an entire demand *system*. In this case, it characterizes demand for the audit firms included in the explicitly modeled choice set plus the amalgam of all other audit firms that represents the "outside good." For example, Guo et al. (2017) include the Big 4 firms along with two medium-size audit firms in the choice set with all other audit firms representing the outside good. Gerakos and Syverson (2015) specify the outside good as all non-Big 4 audit firms. Both studies assume that all clients could have, in principle, chosen the outside good. This assumption might not be valid for large clients. To test this assumption, Gerakos and Syverson (2015) estimated demand for client firms with above-median total assets and found similar parameter estimates. In general demand estimation settings, specifying the outside good typically involves assumptions, because a researcher must consider which other possible choices buyers might have faced, often including the option of choosing not to buy at all. This is made a bit easier in the audit setting for public clients, however, because purchase is mandated.

If we know or can estimate the elements on the right-hand side of Eq. 5, we can compute the utility of each of the choices facing any given audit client. By comparing these utility levels, we can predict client firms' choices for *every* potential audit firm,

---

[7]An indirect utility function substitutes the optimal quantity—which is trivially equal to one in a discrete choice framework (the firm either hires the audit firm or it doesn't)—back into the standard utility function.

not just one at a time.[8] For example, we could compute that a client firm with a particular set of characteristics—assets, foreign sales, etc.—will have a specific level of expected utility from hiring PricewaterhouseCoopers, another expected utility level from hiring Deloitte Touche, and so on. These expected utility levels correspond directly to the probabilities of the client hiring each potential audit firm through the following equation:

$$P_{ij} = \frac{e^{V_{ij}}}{\Sigma_j e^{V_{ij}}}. \tag{6}$$

To go from Eq. 5 to Eq. 6, note that Eq. 5 can be written as $U_{ij} = V_{ij} + \epsilon_{ij}$, in which $V_{ij} \equiv \beta_{ij} x_{ij} - \alpha \ln(p_{ij}) + \delta_j$ is the observable portion of utility, and ϵ_{ij} is the unobserved portion of utility. That, combined with the assumption that ϵ_{ij} is distributed Type I extreme value, implies that the choice probabilities have the characteristic logit form in Eq. 6.[9] Equation 5 is all we need to compute the predicted choice probabilities (which aggregate into market shares) for all of the audit firms in the client's choice set. In other words, Eq. 5 has the demand curves for the choices embodied within it.

Furthermore, computing the change in consumer surplus under policy counterfactuals (i.e., proposed changes to the choice set faced by a client) is particularly easy given the form of Eq. 5.[10] We first compute the utility of the chosen audit firm directly using Eq. 5, which is already net of the price that the client firm pays (and, again, notice that we can use the same equation to do this for each of the audit firm choices—all that changes are the values of the variables on the right-hand side). If under the counterfactual the client's chosen audit firm is to be removed from the client's choice set, we can then use Eq. 5 to compute the utilities of the remaining choices to identify the utility of the client's second best choice. We can then subtract the utility of the chosen audit firm from the utility from the second best choice to calculate the change in utility.

We are almost, but not quite, done at that point, because the change in utility thus computed will be in "utils" rather than dollars. Converting changes in utility to dollars is easy though, because α, the coefficient on price in the utility function, converts units of fees to utils. So, for example, if audit fees enter equation (5) as their logarithm, as is the case in Gerakos and Syverson (2015) and Guo et al. (2017), one just divides the change in consumer surplus computed in utils by α and exponentiates the result to obtain change in consumer surplus expressed in dollars.[11]

---

[8]In their model of lowballing, Kanodia and Mukherji (1994) similarly assume that clients choose audit firms based on client-level utility.

[9]Hence this approach is sometimes referred to as "conditional logit." The benefit of the Type I extreme value distribution is that its integral has a closed-form expression, leading to the straightforward functional form in Eq. 6. By contrast, we could assume that the error terms had some other distribution (multivariate normal for example) but would in this case have to use numerical integration. Assumptions about the ϵ_{ij} distribution do place restrictions on clients' substitution patterns, as we discuss below.

[10]As discussed by Train (2009), the absolute level of utility can be pinned down only to within a constant. Changes in utility can, however, be calculated because the constant drops out.

[11]Similarly, Eq. 5 can be used to compute the expected changes in consumer surplus from entire sets of choices using formulas given by Anderson et al. (1992), among others.

To empirically estimate Eq. 5, one chooses the parameters of Eq. 5 to most closely match the actual choices of client firms in the data. This process involves maximizing the following log likelihood:

$$LL(\alpha, \beta, \delta) = \sum_i \sum_j y_{ij} \ln P_{ij} = \sum_i \sum_j y_{ij} \ln \frac{e^{V_{ij}}}{\Sigma_j e^{V_{ij}}}. \tag{7}$$

Mechanically, this means regressing for every potential client-audit firm pairing an indicator of whether that client actually chose that audit firm in the data (i.e., within each client firm, the indicator will be equal to 1 for one audit firm and 0 for all other audit firms) on all the elements of the right-hand side of Eq. 5: the audit firm's predicted fees and attributes and their interactions with the client's characteristics.

The maximum likelihood estimation of this specification chooses the parameters of utility function in Eq. 5, δ_j, α, and β_{ij}, to fit as closely as possible the choice probabilities predicted by the model to the choices observed in the data. Once we have these parameters, we can compute an audit firm's utility for a particular client in three steps: a) multiplying the estimated parameters by their corresponding observables (e.g., multiplying the estimate of α by the logged p_{ij} observed for the potential client-audit firm pair and multiplying the estimated β_{ij} by their respective client-audit firm x_{ij} values),[12] b) drawing a value for ϵ_{ij} from a Type I extreme value distribution, and then c) summing these products and the ϵ_{ij} draw. This procedure gives us a value U_{ij}, the utility that client firm i receives from hiring audit firm j. Changes in consumer surplus can be estimated by differencing the client's utility for its actual choice from the utility the client would derive if it was forced to choose its next best choice.[13]

Note that there is an "implicit fixed effect" in a conditional logit choice model like Eq. 5. This implicit utility component is a *client-level* fixed effect, not a product-level one. The source of this implicit fixed effect comes from the fact that all utility elements that do not vary across choices (including the outside good) for a given buyer effectively cancel out and therefore drop out of the estimation. Suppose, for example, that there were some direct effect of a client firm's assets on its utility of hiring any audit firm—10 utils per unit of logged assets, just to be specific. Larger client firms would then receive higher utility from their choice. However, because any such effect would add the same amount of utils to the client's utility for all potential audit firms, it would not affect the ordering of the utilities each audit firm would deliver to the client. Because the utility ordering wouldn't be affected, the client's choice wouldn't be affected either.

As an example of this approach, if, all else equal, a lot of clients have hired Ernst & Young, there will be a lot of indicators on the left-hand side of the regression showing that Ernst & Young was a client's observed choice. The estimation procedure will try to match this fact by increasing Ernst & Young's predicted utility level through

---

[12] Because fees are only directly observed in the data for the audit firm that a client actually chooses, predicted fees must be used in the demand estimation for the audit firms that are not chosen. Several different approaches might be used for this, including machine-learning type predictive models. See Gerakos and Syverson (2015) for one example of such an approach and details on its implementation.

[13] For a discussion of how to calculate changes in consumer surplus from discrete choice models, see McFadden (1999).

raising its estimate of Ernst & Young's brand effect parameter—that is, the coefficient on the Ernst & Young fixed effect on the right-hand side.[14] This raises the model's predicted probabilities that Ernst & Young is chosen by clients.

A larger estimate of the coefficient on the brand effect parameter for a particular audit firm indicates that, all else equal, that firm is more likely to be chosen. This is because a high coefficient implies that the firm delivers a high utility, all else equal. Moreover, the brand effects can be cardinally compared across audit firms. Thus the estimated coefficients on the audit firm indicator variables reflect not just the utility that each potential audit firm choice delivers to clients, relative the excluded outside good (which can be thought of as having a brand effect coefficient that is normalized to zero), but also the utility levels that audit firms deliver *relative to each other*.

To see this in an example, let's work off the estimates in Table 7 of Gerakos and Syverson (2015). There, all of the Big 4 indicator variables have positive coefficient estimates. This means that, all else equal, each of the Big 4 is preferred to a non-Big 4 audit firm (which again can be thought of as having a fixed effect coefficient of zero). Now, within the Big 4, Ernst & Young has the largest coefficient estimate, then Price-waterhouseCoopers, then Deloitte, and finally KPMG. These differences indicate that, again all else equal, Ernst & Young is most preferred by clients, then KPMG, then Deloitte Touche, then PricewaterhouseCoopers (though PricewaterhouseCoopers is still preferred to an audit firm outside the Big 4). The magnitudes of these relative preferences are given by the sizes of the coefficients, though these reflect differences in utils and would need to be converted to dollars by dividing through by α. By contrast, Guo et al. (2017) find that the brand effects for the Big 4 in France are significantly negative implying that, all else equal, clients prefer the outside good.

Keep in mind that everything up to this point deals with the Big 4 variables' main effects—that is, their utility effects that are common across all client firms. But the utility specification also includes interactions of client firms' characteristics (e.g., logged assets and segments, foreign sales, etc.) with these indicator variables. What these interactions allow, and what the estimated coefficients on these interactions indicate, is that clients with different characteristics value potential audit firms differently.

Working again off the estimates in Table 7 of Gerakos and Syverson (2015) as an example, the interaction between logged client assets and the Big 4 indicator variables with the largest estimated coefficient is PricewaterhouseCoopers. This means that, as we compare across clients of different asset levels, the relative preference for PricewaterhouseCoopers grows with client size faster than for the other Big 4 audit firms. Thus, while the main utility effect of PricewaterhouseCoopers was smaller than that of the other Big 4 firms, PricewaterhouseCoopers is looked at relatively more favorably by clients with more assets than those with fewer. Thus the total influence of PricewaterhouseCoopers on a client's choice is not just its main effect but also the effect of all the interactions between that client's characteristics and the

---

[14]While often referred to as the "brand effect" in the demand estimation literature, this term captures any choice-specific (here, audit firm-specific) utility component that is common across all potential buyers (here, client firms). This might be, but does need not to be, tied explicitly to the brand itself.

PricewaterhouseCoopers brand effects. We, of course, compute and include all of these elements of brand—main effects and interactions—in our analysis.

The utility function in Eq. 5 also includes price. Here, Gerakos and Syverson (2015) and Guo et al. (2017) follow standard practice in discrete choice demand estimation by imposing that price changes have the same impact on utility for all products; that is α does not vary across client firms. The logic of this standard assumption is straightforward: paying a given amount more in the form of higher prices has the same effect on the buyer's utility, regardless of what choice that expenditure was put toward.

Note, however, that imposing a common price coefficient across all choices does *not* impose that price *elasticities* are the same across all products. In fact, they are not (and only will be in the very special case where choices are observed to have the exact same market shares in the data). In logit demand systems, the price elasticities are a function of both the sensitivity of utility to price changes (α) and the choice probabilities as predicted from all buyer characteristics and product attributes. This relates to the fact that price elasticities are not just about dq/dp (which is closest to α in our case) but also p/q (which in the logit demand system is a function of predicted probabilities). This dependence of the price elasticity on not just α but also all the other components of the utility function explains why there are different own-price elasticities for each audit firm. Cross-price elasticities work in the same way; they depend both on the price sensitivity of utility and the choice probabilities. But now rather than just the probability of choosing one audit firm as in the own-price elasticity case, cross-price elasticities depend on the probabilities of two choices.

Demand estimation does face a conceptual problem related to the aforementioned confounding of supply and demand influences in the audit fee regression. Namely, we might expect that prices are correlated with unobserved demand differences, because firms will take demand shifts into account when setting prices. For example, if an audit firm recognizes that certain kinds of potential clients have a higher willingness to pay for its audits, it might well quote such firms higher fees. We would then observe in the data that these clients hire the audit firm despite seemingly high fees. A model that simply tries to fit the observed correlation between purchase probability and fee levels would therefore deliver a positively biased estimate of the price responsiveness coefficient α. In other words, demand would be estimated to be less responsive to prices than it truly is. Essentially, the model would be incorrectly explaining those firms' willingness to pay high fees to that particular audit firm as reflecting insensitivity to fees, in general. That is, it would be confusing a shift in demand for a change in the slope of the demand curve.

Instrumental variables methods are again the solution to this problem. An instrument (a factor shifting the supply curve but not the demand curve) allows one to obtain unbiased estimate of the price sensitivity of demand. Gerakos and Syverson (2015) use the unexpected implosion of Arthur Andersen as an instrument, because it created an exogenous shift in the supply of audit services. We examined how client firms reacted to fee changes *driven specifically by this supply shift*, as implemented through an instrumental variables regression, to estimate the price sensitivity of demand. Other demand estimation studies have used different instruments, and the

best in any given setting will depend on the specifics of the market, but the principle behind all these alternatives is the same.

Another conceptual issue involved with conditional logit approaches regards the structure they implicitly assume about buyers' substitution patterns. As discussed by Gerakos and Syverson (2015), specifying that the error terms are IID, as the logit model does, imposes the independence of irrelevant alternatives (IIA) assumption. Namely, for any two potential choices *a* and *b*, the relative probability of a buyer choosing *a* over *b* does not change as alternative choices are added to, or removed from, the buyer's choice set. (This property is often described using the classic "Red Bus, Blue Bus" example.) This property does not, however, appear to be a major concern in the audit setting. Consistent with the IIA property, Gerakos and Syverson (2015) show that relative market shares were relatively stable before and after the disappearance of Arthur Andersen.

4 How competitive is the audit market?

The level of competition in the market for audits of public companies is a first-order question in the audit literature. It is a critical input into policy discussions about whether the audit market is too concentrated, the costs and benefits of mandatory audit firm rotation, and the extent to which the level of competition affects the quality of audits, among others. Furthermore, understanding the level and nature of competition is necessary to interpret empirical studies of the audit market more generally.

The demand estimation techniques described above can offer quantitatively sharp answers to the competitiveness question. We explain how in this section.

Some researchers conjecture that the audit market is highly competitive, possibly even perfectly competitive within the Big 4 and non-Big 4 segments.[15] Perfect competition has several related implications for the audit market. One, new firms could easily enter the market; that is there would be free entry. For example, after the disappearance of Arthur Andersen, an alternative supplier could have entered the market. Two, clients would view audit firms as perfect substitutes. Audit firms could be randomly reassigned to clients, and the clients would be indifferent. This condition would also exclude the possibility of either switching costs or heterogeneity in the services provided by audit firms (e.g., industry specialization). Three, the residual elasticity of demand faced by an audit firm would be negative infinity. That is, if an audit firm were to raise fees infinitesimally, it would lose all of its clients.

Demand estimation methods deliver an estimate of the elasticity of demand faced by sellers. Demand elasticity is a theoretically founded indicator of the extent of competition in an industry. It is also one of the most commonly applied empirical measures of competition.[16] The elasticity of demand for an audit firm measures the

---

[15] Examples include GAO (2008) and Dunn et al. (2011), and Dunn et al. (2013).

[16] While concentration measures (like the Herfindahl-Hirschman index) are sometimes used for measuring the extent of competition, high concentration can occur in both highly competitive and highly uncompetitive industries, as discussed by Sutton (1991) and others.

percentage change in the probability that a client will choose the audit firm for a one percent increase in the price charged by the audit firm.

Assuming that audit firms maximize profit when setting prices, one can use estimates of the demand elasticity, ϵ_D, to estimate price-cost margins using the well-known expression for a firm's profit-maximizing price:

$$\frac{P}{MC} = \frac{|\epsilon_D|}{|\epsilon_D| - 1}, \tag{8}$$

in which P is price and MC represents marginal cost. If the market is perfectly competitive (i.e., clients leave if the audit firm raises price even slightly), then $\epsilon_D = -\infty$, leading to a ratio of price to marginal cost of one (i.e., $P = MC$). As demand becomes more inelastic (i.e., moves toward zero), margins increase.

In Panel B of Table A.1 of Gerakos and Syverson (2015), we provide estimates of demand elasticities for former Arthur Andersen clients. The implosion of Arthur Andersen provides a setting to estimate long-run demand elasticities that are not affected by switching costs. For these clients, we find elasticities in the neighborhood of -1.6, which is sufficiently far from negative infinity to rule out that the audit market is perfectly competitive (or even close to being perfectly competitive). These elasticity estimates imply an optimal markup of $P/MC = 2.67$, both statistically and economically well above the baseline of one under perfect competition. Simply put, clients' demand for audit services, *as revealed by their own behavior*, indicates they view audit firms as being far from perfect substitutes.[17]

Substantial price-cost margins do not, however, necessarily imply that audit firms earn substantial bottom line rents. Firms can earn substantial price-cost margins yet dissipate them via nonprice competition. Industries can appear competitive while not be being price rivalrous—Coke versus Pepsi being a classic example. Competition can be about frittering away rents from market power (Posner 2001), and there can be a lot of "head butting" in a market even if markups are large. Consider, for example, residential real estate agents. Markups are high in this industry, yet agents burn up the markups by spending time and money competing to obtain clients (Hsieh and Moretti 2003). It is important to point out that such non-price competition is inconsistent with sellers providing undifferentiated products, as would be implied by perfect competition.[18] Consistent with audit firms providing differentiated services, there is a large literature on industry specialization by audit firms. Moreover, Brown and Knechel (2016) find evidence that clients choose audit firms based on audit firm characteristics in addition to industry specialization. These results provide evidence against perfect competition. With respect to the competitiveness of the market, the discrete

---

[17] We find non-Andersen clients exhibit even less elastic demand for the audit firm that they had hired in the prior year, on the order of -0.3 (Panel C of Table 7). However, these are short-run elasticities that audit firms are unlikely to use as pricing guides.

[18] The industrial organization and marketing literatures typically divide product differentiation into vertical and horizontal dimensions. Under vertical differentiation, all market participants share the same rankings of products' quality levels. Under horizontal differentiation, competing products differ in their characteristics and consumers differ in their evaluation of the product characteristics. Differentiation of either type can confer market power to a seller.

choice demand approach provides estimates of the residual elasticity of demand, which speaks to the extent of imperfect competition.

5 Going forward

We argue that researchers in the empirical auditing literature can make progress by applying commonly used techniques in the industrial organization and quantitative marketing fields. Both of these fields address questions similar to those in the auditing literature (e.g., the level of competition in a market, the relationship between competition in a market and the quality of goods and services offered in the market, the effect of regulation on the level of competition, and the dynamics of pricing), and they have developed empirical methods to address these questions. Yet there is at present a disconnect between the auditing literature and these literatures. One manifestation of this is an overreliance on hedonic price regressions. By contrast, (Guo et al. 2017) provide an excellent example of applying techniques from industrial organization and quantitative marketing to address an important question in the auditing literature that cannot be addressed by estimating an audit fee regression. We encourage more such studies.

References

Anderson, S., De Palma, A., & Thisse, J. (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge: The MIT Press.

Angrist, J., & Krueger, A. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(1), 69–85.

Brown, S., & Knechel, W. (2016). Auditor-client compatibility and audit firm selection. *Journal of Accounting Research*, *54*(3), 725–775.

Copley, P., Gaver, J., & Gaver, K. (1995). Simultaneous estimation of supply and demand of differentiated audits: Evidence from the municipal audit market. *Journal of Accounting Research*, *33*(1), 137–155.

Deis, D., & Hill, R. (1998). An application of the bootstrap method to the simultaneous equations model of the demand and supply of audit services. *Contemporary Accounting Research*, *15*(1), 83–99.

Dunn, K., Kohlbeck, M., & Mayhew, B. (2011). The impact of the Big 4 consolidation on audit market share equality. *Auditing: A Journal of Practice and Theory*, *30*(1), 49–73.

Dunn, K., Kohlbeck, M., & Mayhew, B. (2013). *The impact of market structure on audit price and quality, Unpublished working paper* . Madison: University of Wisconsin.

GAO (2008). Audits of public companies: Continued concentration in audit market for large public companies does not call for immediate action. Tech. Rep. GAO-08-163, United States Government Accountability Office.

Gaver, J., & Gaver, K. (1995). Simultaneous Estimation of the Demand and Supply of Differentiated Audits. *Review of Quantitative Finance and Accounting*, *5*(1), 55–70.

Gerakos, J., & Syverson, C. (2015). Competition in the audit market: Policy implications. *Journal of Accounting Research*, *53*(4), 725–775.

Guo, Q., Koch, C., & Zhu, A. (2017). Joint audit, audit market structure, and consumer surplus. Review of Accounting Studies, this issue.

Hay, R., Knechel, W., & Wong, N. (2006). Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research*, *23*(1), 141–191.

Hsieh, C.H., & Moretti, E. (2003). Can free entry be inefficient? Fixed commisssions and social waste in the real estate industry. *Journal of Political Economy*, *111*(5), 1076–1122.

Kanodia, C., & Mukherji, A. (1994). Audit pricing, lowballing and auditor turnover: A dynamic analysis. *The Accounting Review*, *69*(4), 593–615.

McFadden, D. (1999). Computing willingess to pay in random utility models. *Trade, Theory, and Econometrics: Essays in honour of John S Chipman*, *15*, 253–274.

Posner, R. (2001). *Antitrust Law*, 2nd edn. Chicago: University of Chicago Press.

Rosen, S. (1974). Hedonic prices and implict markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Simunic, D. (1980). The pricing of audit services: Theory and evidence. *Journal of Accounting Research*, *18*(1), 161–190.

Sutton, J. (1991). *Sunk Costs and Market Structure: Price Competition, Advertising and the Evolution of Concentration*. Cambridge: The MIT Press.

Train, K. (2009). *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.