Economies of Scale in the Audit Market: A Nonparametric Approach

Anne Albrecht*
Texas Christian University
a.albrecht@tcu.edu

Stephanie Miller
Quinnipiac University
sdmiller@qu.edu

Frank Murphy
University of Connecticut
francis.murphy@uconn.edu

*Corresponding Author

February 2022

Abstract: This paper examines economies of scale in the audit market. We use a measure of economies of scale that corresponds to the theoretical framework from the economics literature. Using kernel-weighted local regression analysis, we document evidence of economies of scale among both Big-4 and non-Big-4 auditors and at the city-industry, national-industry, and city (irrespective of industry) levels. By using nonparametric techniques, we are able to identify the audit firm size at which average costs are minimized. We conclude that most audit firms or offices can achieve economies of scale with fewer than ten clients. In addition, we revisit prior literature and suggest that the inability to document scale economies in broad settings may be a byproduct of using linear regression to identify a nonlinear relation.

JEL Classifications: C14, H25, and M42

Keywords: Nonparametric Methods, Audit Fees, Economies of Scale

This work was supported by: Texas Christian University, the UConn School of Business, Quinnipiac University, and Yale University. This study has benefited from the thoughtful comments from Margaret Banker, Jeremy Douthit, Jacob Jaggi, Stephen Lusch, Elaine Mauldin, Devan Mescal, Paul Michas, Karen Nelson, Sarah Parsons, Dan Russomanno, Steve Utke, Tara Vakil, Robert Whited, Sunny Yang, Spencer Young, Tiemen Woutersen, Nina Xu the participants at the 2018 Northeast Regional Meeting, 2019 Hawaii Accounting Research Conference, 2021 AAA annual meeting, University of Connecticut and Texas Christian University. Frank Murphy would like to acknowledge the Nicholas DuBiago '66 Accounting Fund for providing research support

I. INTRODUCTION

Economies of scale represent cost advantages that arise when increased output leads to reduced marginal costs of producing a good or service. After reaching a minimum average cost, further expansion creates diseconomies of scale, resulting in a U-shaped average cost curve. Prior research on audit markets finds evidence of industry-specific economies of scale among Big-4 audit offices (i.e., at the city level) (Cahan, Jeter, and Naiker 2011; Fung, Gul, and Krishnan 2012; Bills, Jeter, and Stein 2015) but fails to find similar evidence at the overall city level (non-industry specific), national level, or outside of Big-4 firms. However, economic theory does not suggest that economies of scale occur only within industries, for larger audit firms, or at the office level. Given this mismatch between theory and prior findings, we revisit economies of scale in the audit market using nonparametric techniques to examine a broad range of settings in which they are expected to arise.

Audit firms realize economies of scale through investments in the resources, processes, training, technology, and shared services required to perform audits that are standardized across clients. Theoretically, economies of scale decrease the average cost of producing audit services as firms take on more clients or as their current clients grow.[1] These decreases in average cost may manifest from the spreading of fixed costs over additional units of output or from efficiencies gains in performing procedures. For example, audit firms often standardize the process for auditing accounts receivable. As an audit firm takes on additional clients or as its existing clients grow, it will audit additional accounts receivable balances using the same processes. While larger clients require more audit effort and thus are associated with higher costs and audit fees, we expect audit effort increases at a decreasing rate due to economies of scale.

[1] Client growth may alternatively result in additional audit risk, increasing costs.

The current audit literature focuses on economies of scale that arise in Big-4 audit firms because of industry specialization. These studies theorize that industry specialization leads to greater homogeneity in required audit inputs. Accordingly, audit firms realize cost savings as they perform similar audit tasks across clients that operate in the same industry. While several studies have found results in line with this notion (e.g., Fung, Gul, and Krishnan 2012), we expect economies of scale to also arise in broader circumstances not necessarily related to Big-4 status or industry expertise. Several audit processes, such as those related budgeting, risk assessment, and testing common financial statement line items, apply to clients in all industries and are standardized to streamline the process without diminishing quality. Shared services such as general counsel, human resources, information technology, and specialists normally support all audits independent of industry. Similarly, audit-specific technology and data analytics tools developed to assess risk and perform audit procedures apply to clients across industries. These sources of potential efficiencies and declining average costs are applicable to both Big-4 and non-Big-4 auditors. Thus, there are a broad range of settings in the audit market that could give rise to economies of scale. Consistent with this intuition, Gong, Li, Lin, and Wu (2016) find efficiencies when Chinese audit firms merge. We expand this literature by examining economies of scale in settings beyond Big-4 industry specialists.

We measure of economies of scale using the average audit fee per million dollars of client assets.[2] Our measure of economies of scale reflects the average cost per unit of output, resulting in a homogenous measure of unit cost that we can use to compare across diverse audit firms, clients, and industries. We use assets audited to measure the audit firm's unit of output because prior research finds that assets explain approximately 70 percent of the variation in audit fees (Hay,

[2] Because cost (e.g., billable hours) is unobservable, audit fees proxy for audit cost to the extent that cost savings are passed on to clients.

Knechel and Wong 2006), suggesting assets reflect the magnitude of the completed audit (i.e., output). Additionally, as an average cost measure, our measure more closely conforms to economic theory on economies of scale than measures from prior research that use auditor size as a proxy for scale. We nonparametrically examine how this average cost changes across the distribution of audit firms' number of public clients.[3] We use nonparametric analysis because the classical U-shaped theoretical model of economies of scale suggests a nonlinear relation between scale and cost. While OLS regression is standard in audit research, the ability to identify a theoretically nonlinear relation by fitting a straight line to the data is significantly limited. In contrast, nonparametric estimation fits a curve along the entire distribution. By providing evidence beyond average relations, nonparametric analysis allows us to identify the point at which costs are minimized. Nonparametric analysis provides a range of statistical techniques that do not require the restrictive assumptions of OLS, resulting in a much richer set of information when visualizing associations between variables.[4] The reliance on OLS and lack of consideration of nonlinearity in prior research may hinder the identification of economies of scale outside of Big-4, industry-specialist offices.

Our analyses primarily use kernel-weighted local linear regression, a nonparametric technique that uses a series of localized regressions, weighting observations that are close to one another more heavily to produce a regression output that is not necessarily linear. The output of this analysis is a graph that enables us to determine where along the curve average cost is minimized. We perform this nonparametric estimation at the city-industry level, the national-

[3] Consistent with prior research, we are unable to observe the audit firm's portfolio of nonpublic clients. Therefore, our analysis is limited to public clients.
[4] For example, these tools do not require the assumption that the relation between variables is linear or that the researcher knows the underlying probability distribution (e.g., normal, gamma, etc.).

industry level, and the city level for both Big-4 and non-Big-4 auditors.[5] Because of the competitive divide between Big-4 and non-Big-4 auditors, we expect average costs to be different among Big-4 auditors compared to non-Big-4 auditors. Therefore, we examine Big-4 and non-Big-4 auditors separately, consistent with prior research. We do not examine the cross-sectional national level regardless of industry because of the limited number of Big-4 observations (i.e., one per year).

In each setting, we find economies of scale and note that cost savings arise fairly quickly (i.e., close to the left side of the distribution). After this initial downward slope, average fees level off in most specifications. In contrast, we find limited evidence of diseconomies of scale, which is unsurprising as accepting new audit clients is an endogenous decision, and on average, we do not expect audit firms to take on clients that lead to inefficiencies. Overall, our results suggest that both Big-4 and non-Big-4 auditors realize economies of scale from taking on additional clients, although these scale savings are more prominent for non-Big-4 auditors. The number of public clients at which average fees are minimized differs by audit firm type, but the cost functions of Big-4 and non-Big-4 auditors appear to converge. In most settings, we find evidence of economies of scale with fewer than 10 public clients, and the marginal savings from taking on an additional public client significantly decreases starting between four (city level) and 20 (national-industry level) public clients. Despite this finding, most small audit firms audit very few public clients and thus may not realize economies of scale cost savings.

Our findings support the theorized nonlinear economies of scale, the broader settings in which they can arise, and the use of nonparametric analysis. However, because prior research documents economies of scale only at the city-industry level for Big-4 auditors, we revisit those

[5] In untabulated analysis, our inferences are not sensitive to changes in the bandwidth or using kernel-weighted local quadratic or cubic regression.

findings. Fung et al. (2012) use the interaction between audit office scale (based on number of clients in a city-industry) and audit office industry specialization (based on the office with the greatest total fees in a city-industry) as a proxy for city-industry-level economies of scale. They find a negative relation between this interaction term and audit fees, which they attribute to industry-specialist auditors passing on cost savings generated through economies of scale to clients.

Within their sample period (2002-2007), we successfully replicate the results of Fung et al. (2012). When we extend the sample period to 2018, we find that audit office scale is negatively associated with audit fees for firms that are not industry specialists as well. This result suggests that even without industry specialization, large audit firms are able to pass on some economies of scale savings to clients, consistent with the findings of our nonparametric analysis. We are cautious about interpreting this result, however, because it is confined to the extended sample period and uses OLS to examine a nonlinear relation. We consider nonlinearity by adding nonlinear terms to the OLS regression models as suggested by Shipman, Swanquist, and Whited (2017). When higher-order terms are included in the primary specification, nearly all their coefficients are statistically significant, providing additional evidence that audit fees have a nonlinear relation to audit office size. Interestingly, when nonlinear terms are included in the model, audit office scale is significant, but the interaction between audit office scale and industry specialization is no longer significant. This finding suggests that identification of economies of scale at the city and/or city-industry levels depends on model specification, and thus the functional form of this relation requires additional consideration.

Further, Whited, Swanquist, Shipman and Moon (2022) note econometric issues when several proxies capture a similar construct, which they refer to as "same construct controls." In the

Fung et al. (2012) model, the variables measuring audit office scale, industry specialization, and city size all capture audit office size to some extent. We find evidence that these "same construct controls" bias the coefficient estimates, consistent with functional form misspecification and underscoring the importance of using nonparametric estimation techniques in identifying economies of scale. Importantly, our evidence also suggests that the inability of prior research to document scale discounts outside of Big-4 industry specialists is likely attributable to research design choices rather than the absence of scale economies.

Our paper makes two key contributions. First, we add to the literature on economies of scale in the audit market. While economies of scale seem to be a forgone conclusion, prior studies have not been able to document their existence beyond limited settings. These studies use measures of audit office size based on magnitude or decile-ranked variables to capture scale. Our measure, audit fees per million dollars of client assets, captures average cost and thus more closely conforms to the theoretical model of economies of scale. Additionally, our measure is simple to construct and straightforward to interpret. Using this measure, we document the existence of economies of scale among Big-4 and non-Big-4 audit firms, regardless of industry expertise. More importantly, our paper is the first to document the audit firm/office size at which economies of scale are realized, generating normative suggestions on the level of output at which an audit firm may expect average costs to be minimized. Thus, our analysis informs regulators and audit firms on optimal market structure. We find that economies of scale can be realized with as few as four public clients.

Second, our study contributes to the methodological literature in accounting. Our analyses raise questions about the functional form and broader empirical approach employed in the existing auditing literature. We use nonparametric analysis, a set of tools underutilized in accounting research in general that can act as a powerful complement to OLS. Nonparametric analysis allows us to better document nonlinear relations, to analyze full distributions, and to study variables of interest more broadly than prior research. As a result, we answer the call from Gow, Larcker, and Reiss (2016) for more studies in accounting research that provide a rigorous understanding of the underlying relation between variables of interest.

II. BACKGROUND

As depicted in Figure 1, the classical model of economies of scale theorizes a nonlinear average cost curve. As a company increases its level of output, marginal costs decrease at a decreasing rate (economies of scale) and eventually begin to increase again at an increasing rate (diseconomies of scale). The resulting long-run cost curve includes an optimal company size at which average unit costs are minimized.

[Insert Figure 1 here]

Economies of scale are often the result of fixed costs being spread across additional units of output or improved efficiencies as production increases, both of which should be applicable to audit firms. For example, Stigler (1958) notes that research and technological innovation are complex, require significant investment, and may influence economies of scale. Audit firms make infrastructure investments that are likely independent of client characteristics. This infrastructure may generate economies of scale through technological efficiency, the spreading of training and other fixed costs, the ability to draw from a database of knowledge held at the firm level, reduced advertising costs due to reputation, and so on. For example, employing a tax specialist who audits multiple clients' income tax provisions not only spreads the direct cost (e.g., salary) over a number of clients but also leads to efficiencies from applying the knowledge of common tax law changes to multiple clients. Audit firms should also generate efficiencies and realize economies of scale by applying similar audit processes across multiple clients or offices. For example, the processes for assessing risk; determining materiality; budgeting; planning; and auditing common, less complex financial line items should be similar across diverse clients. Additionally, auditors are incentivized to standardize their processes to produce, monitor, and replicate high-quality audits (i.e., output).

Economies of scale arise when output increases. In the audit market, the quantity of output depends on both the number and size of clients audited. Audit firms may realize cost savings on large clients in particular if these audit engagements involve many similar procedures that can be standardized and repeated throughout the engagement. Alternatively, if, as client size increases, there is an increasing chance that the audit becomes "self-contained," with audit team members working on only a single client, this could limit the ability to generate spillover effects that can lead to economies of scale across multiple clients. Heterogeneity in client size and other characteristics makes it difficult to assess whether economies of scale are observed in the audit market, as no single variable or coefficient represents average unit cost.

Using meta-analysis Hay, Knechel, and Wong (2006) document that some measure of client size is nearly always present when estimating the empirical audit fee model and that it is usually positive and statistically significant. Client size, commonly measured as the natural logarithm of total assets, often explains over 70 percent of the variation in audit fees.[6] This relation could be due at least to two direct channels: effort and risk. The literature attempts to disentangle these effects by including proxies for complexity (e.g., number of subsidiaries, foreign operations, extraordinary items, discontinued activities), which is positively related to audit fees (e.g., Hay 2013). Client riskiness is also associated with higher audit fees (e.g., Schelleman and Knechel

[6] Sales is another common measure for client size.

2010). Large clients, while often more complex, may be more likely to have robust internal controls, reducing risk (Hogan and Wilkins 2008). Thus, audit fees may increase at a decreasing rate with client size to the extent that large companies are less risky or at an increasing rate to the extent that size also captures complexity.

Prior audit studies focus on the relation between audit fees and audit firm size, arguing that a positive relation suggests demand for high-quality audits and a negative relation suggests economies of scale. Simunic (1980) finds mixed evidence supporting both scale economies and premiums for audit quality. This is the only study to document economies of scale in a broad setting. Instead, the literature has subsequently extensively documented the Big-N price premium, suggesting that Big-N audit firms are able to extract higher fees from clients due to market dominance, reputation effects, and the provision of high-quality audit services.[7] Francis and Stokes (1986) observe a Big-N price premium for small clients but not for large clients in the Australian audit market. They conclude that higher Big-N prices for large clients are obscured by higher non-Big-N prices that arise from diseconomies of scale when small audit firms audit large clients. The distinction between Big-N and non-Big-N audit firms is a national-level characteristic. Because local audit offices are the primary decision units, recent research focuses on audit office size. Consistent with size being a proxy for quality, Choi et al. (2010) find a positive relation between audit fees and audit office size. Francis, Mehta, and Zhao (2017) find audit offices that gain (lose) a major industry client increase (decrease) audit fees for other clients in the same industry.[8]

[7] See Francis (1984); Palmrose (1986); Francis and Stokes (1986); Francis and Simon (1987); Beatty (1993); Chan, Ezzamel, and Gwilliam (1993); Craswell, Francis, and Taylor (1995); DeFond, Francis, and Wang (2000); Chaney, Jeter, and Shivakumar (2004); McMeeking, Peasnell, and Pope (2006); Clatworthy, Makepeace, and Peel (2009); Hay (2013); Lennox, Francis, and Wang (2012), among others.

[8] In a related stream of literature, there is debate over functional form and whether client characteristics drive the Big-N quality effect (e.g., Lawrence, Minutti-Meza, and Zhang 2011; DeFond, Erkins, and Zhang 2016).

To the extent that each audit engagement represents a unique product, industry specialization may offer greater opportunities to capitalize on economies of scale. Hay (2013) notes "specialization at the local level implies that knowledge is held by individuals or the audit team; specialization at the national level implies that the benefits come from a database of knowledge held by the firm as a whole." Mayhew and Wilkins (2003) argue that industry specialization decreases auditor cost by allowing auditors to both spread industry-specific training across more clients and more efficiently address industry-specific issues. The auditor industry specialization literature largely finds a fee premium for specialist auditors, indicating higher audit quality among industry specialists (Francis, Reichelt, and Wang 2005; Basioudis and Francis 2007; Zerni 2012; Hay 2013), although there is debate as to whether this specialist premium arises at the city, national, or global level.

The economics literature identifies various types of expansion that could generate cost savings, such as increased production of a single product (economies of scale), production of multiple products with shared inputs (economies of scope), and increasing market share (economies of density) (Panzar and Willig 1981; Baumol, Panzar, and Willig 1986; Caves, Christensen, and Tretheway 1984). The audit literature generally refers to all three of these types of cost savings as economies of scale. Current research (e.g., Choi et al. 2010, Fung et al. 2012, Bills et al. 2015) captures economies of scope by examining audit firms' number of clients (insofar as each audit represents a unique product) and economies of density by examining their market share (e.g., DeFond, Francis, and Wong 2000; Mayhew and Wilkins 2003; Pearson and Trompeter 1994). In contrast, we focus on cost per unit of output in the audit market to better capture economies of scale as measured in the economics literature and underlying theoretical model. For the remainder of this paper, we do not differentiate between economies of scale, scope, and density and simply refer to all of these as economies of scale for consistency with prior audit literature.

Our study is closely related to Fung et al. (2012), who find evidence of both industry-specialization premiums and scale discounts using a sample of companies audited by Big-4 auditors. We extend their findings in three important ways. First, Fung et al. (2012) examine audit office size as the unit of measurement. Instead, we examine the cost function of the audit office using a measure of cost per unit of output that is more homogenous (i.e., akin to a production function of a commodity). Second, we utilize nonparametric analysis to alleviate concerns over functional form misspecification and to identify the point along the distribution at which economies of scale arise. Third, we investigate economies of scale more broadly, examining both Big-4 and non-Big-4 audit firms at the city, national, and industry-specialization levels, as theory suggests city-level industry specialization is just one of many ways firms can generate scale-related cost savings.

III. RESEARCH DESIGN AND SAMPLE SELECTION

Nonparametric Analysis

In this section, we provide a brief overview of nonparametric analysis and how it can be used to examine economies of scale. We present additional econometric discussion and illustrative examples in Appendix B. For additional discussion, see Härdle and Linton (1994) or Cameron and Trivedi (2005).

OLS regression is appealing because it is well known and straightforward to implement. Under certain assumptions, it generates the best (in terms of minimum variance) linear unbiased coefficient estimates. However, OLS is not appropriate for all empirical analyses. For example, when the association between two variables is nonlinear, the use of a linear estimator can produce a high frequency of large errors, resulting in an inferior estimator compared to nonparametric techniques (Kennedy 2008). As demonstrated in Figure 1, economies of scale are identified based on nonlinear cost curves. This distribution suggests that using OLS to estimate a univariate regression of cost on quantity of output would result in a poor fit. To the extent that both economies of scale and diseconomies of scale exist, it is likely that using OLS would result in a coefficient estimate close to zero and statistically insignificant. Further, OLS yields only average associations between variables. While OLS can be used to document the existence of economies of scale on average, firms cannot extrapolate the point estimates from OLS to approximate the number of clients at which costs are minimized based on their particular facts and circumstances (e.g., a small audit firm with two public clients might receive a larger marginal benefit from adding a client than the average Big-4 auditor). Thus, the use of OLS is not appropriate as it is unable to detect the U-shaped average cost function.

Nonparametric techniques were pioneered by Rosenblatt (1956), Nadaraya (1964), and Watson (1964), among others, and can be used to present rich depictions of data and regression analysis. In this paper, we employ kernel-weighted local linear regression to perform univariate regressions. Locally weighted regression estimates a local slope coefficient by taking the average values of the dependent variable for points that are within some specified distance (bandwidth) of each observation of the independent variable. Using a kernel weighting function, we can obtain a weighted estimate of the local slope coefficient. These weights can be either constants (Opsomer and Breidt 2011), linear, or higher-order (Fan 1992).[9]

[9] Higher order polynomials can be estimated in a similar fashion. See Härdle and Linton (1994) equation (21) or Cameron and Trivedi (2005) equation (9.31). See Fan and Gijbels (1996) for properties of these higher order polynomial estimators. It is important to note that kernel density regression does not handle end points as well as locally weighted linear regression (Fan 1992).

There are two primary research design choices in locally weighted regression: bandwidth and weighting function. Bandwidth is the distance around data points that will be evaluated by the weighting function and is analogous to bin width in a histogram. Jones, Marron, and Sheather (1996) describe the importance of bandwidth choice as follows: "when insufficient smoothing is done, the resulting density or regression estimate is too rough and contains spurious features that are artifacts of the sampling process. When excessive smoothing is done, important features of the underlying structure are smoothed away" (p. 401). Therefore, a researcher must select an appropriate bandwidth so as to not under- or over-smooth the data. As a best practice, authors should ensure results are robust to alternative bandwidths (Cameron and Trivedi 2005).

The weighting function specifies how much weight to assign to each observation within the bandwidth based on the premise that points closer to one another are likely more representative and should be given more weight than points further from one another. In general, the choice of bandwidth is more important than the choice of weighting function, because many weighting functions have similar statistical properties. For our analysis, we use the Epanechnikov kernel weighting function and optimal (default) bandwidth to perform locally weighted linear regression.[10]

Economies of Scale Measure

In this section, we create a novel measure of economies of scale that corresponds to the theoretical construct and captures factors related to both the size and number of audit clients. As discussed in Section II, while economies of scale could certainly be present at the city-industry level, there may be broader channels through which audit firms realize economies of scale.

[10] The Epanechnikov kernel weighting function is generally considered the most efficient and thus is the default option in Stata.

Because economic theory suggests economies of scale may arise at the city or national level and within or across industries, we also construct our variable at the city level (irrespective of industry) and the (national) industry level.

To measure economies of scale, we construct *Fees per Asset* by dividing the sum of audit fees at the audit-firm level in year *t* by the sum of clients' assets (in millions) in year *t*. We then examine economies of scale within city-industries, national industries, and cities irrespective of industry. This creates measures of the average fee per million dollars of assets audited by each firm in each city-industry-year, national-industry-year, and city-year combination. Our measure of economies of scale reflects the average cost per unit of output, resulting in a homogenous measure of unit cost that we can use to compare across diverse audit firms, clients, and industries. We use assets audited to measure the audit firm's output because prior research finds assets explain approximately 70 percent of the variation in audit fees (Hay, Knechel and Wong 2006), suggesting assets capture the magnitude of the audit (i.e., output). We do not argue that each dollar of assets is the same or requires the same audit effort. Rather we explicitly assume that, on average, assets represent a standard unit of measurement that reflects the magnitude of the completed audit (i.e. output). We use *NClients,* the number of unique clients for an audit firm at either the city-industry, national-industry, or city level to capture scale of the client base.

Sample Selection

We begin with the total population of U.S. domiciled companies between 2002 and 2018 from the cross-section of Compustat and Audit Analytics. We remove observations with missing total assets or audit fees. This results in a sample of 90,849 company-years. While many studies exclude the financial services industry due to its different regulatory environment, we choose to retain these observations in our sample.[11] Because our analysis focuses on univariate regression, it does not suffer from sample attrition due to the structurally different balance sheets of financial firms. We then aggregate observations to the audit firm level within city-industries, national industries, and overall cities. Consistent with prior research, we define industries based on two-digit SIC codes (e.g., Francis, Reichelt, and Wang 2005; Reichelt and Wang 2010). This results in 55,943 city-industry audit firm observations, 21,305 national-industry audit firm observations, and 15,181 city-level audit firm observations. Our sample includes the most observations at the city-industry level and the least observations at the city level because an audit firm can operate in multiple industries within a single city.

Descriptive Statistics

In Table 1, Panel A we present descriptive statistics for our economies of scale variable in each audit-firm-level setting. Consistent with city-industry being the narrowest setting we investigate, the average number of clients audited is the smallest in this setting at approximately two clients. We also find that *Fees per Asset* are lowest at the city-industry level, suggesting economies of scale may be most relevant at this level. The distribution of the number of public audit clients reflects several small audit firms that audit only one or two public clients in a given city-industry. At the broader national-industry and city levels, the average number of clients per audit firm is approximately four and six, respectively. Interestingly, average *Fees per Asset* are highest at the national-industry level, more than double the average at the city-industry and city levels. This likely reflects the difficult nature of achieving national-level, industry-specific

[11] Our results are robust to removing observations in the financial services industry.

economies of scale across geographically dispersed offices, where knowledge spillovers are less likely to occur.

Table 1, Panel B presents descriptive statistics for both *Fees per Asset* and *NClients* by Big-4 status. Unsurprisingly, Big-4 auditors appear to be more cost-efficient and audit more public clients than non-Big-4 auditors. Specifically, we find that, at all levels, Big-4 auditors have lower *Fees per Asset* and greater *NClients* than non-Big-4 auditors. This is likely partly due to the fixed cost associated with conducting an audit even for small clients. This highlights the importance of evaluating Big-4 and non-Big-4 auditors separately in our analyses.

[Insert Table 1 Here]

To provide descriptive evidence on our measure of economies of scale, we examine the mean difference in *Fees per Asset* relative to net increases and net decreases in an audit firm's public client base. If economies of scale are present, we expect *Fees per Asset* to decrease (increase) with net increases (decreases) in the number of clients audited. The results of this analysis are presented in Table 2. In Panel A, we limit our samples to audit-firm observations that experience a net increase in the number of public clients in year *t*. We compare the mean *Fees per Asset* in year *t-2* to the mean *Fees per Asset* in year *t*.[12] We find that *Fees per Asset* are significantly lower in the year of a net increase to an audit firm' public client base at both the city-industry and city levels for Big-4 auditors. For non-Big-4 auditors, we find *Fees per Asset* are significantly lower in the year of a net increase to a firm's public client base at the city-industry level. These results are consistent with audit firms passing on the savings from economies of scale to their audit clients. We do not find a significant decrease in average *Fees per Asset* with increases in public

[12] We use a two-year window to avoid fluctuations in audit fees arising from mid-year auditor changes.

client base at the national-industry level. Similarly, in Panel B of Table 2, we examine audit-firm observations with net decreases in the number of public clients in year *t*. As expected, we find that *Fees per Asset* are significantly higher in the year following a net decrease to a firm's public client base in all audit firm specifications regardless of Big-4 status.[13] Taken together, these results provide initial evidence of economies of scale in the audit market.[14]

[Insert Table 2 Here]

IV. DETECTING ECONOMIES OF SCALE

City-Industry Level

We begin our nonparametric analysis at the city-industry level using locally weighted regression on observations up to the 99th percentile of number of clients. The distribution is not well defined in the far-right tail, and thus nonparametric analysis becomes less informative. Our results, presented in Figure 2, provide strong evidence of economies of scale at the city-industry level for both Big-4 (Panel A) and non-Big-4 auditors (Panel B). We present results for these groups separately since the scale of the average fee varies greatly across the two groups. In both groups, we observe a decreasing average fee per million dollars in assets as audit firms in a given city-industry-year take on additional clients. For Big-4 auditors, average fees appear to reach their minimum around four clients in a city-industry-year. The average cost curve remains relatively flat beyond this point. Non-Big-4 auditors' average fees per assets declines slightly after four clients, though it appears to be flattening as well. The flatter cost curve for Big-4 auditors compared

[13] In untabulated results, we examine net increases and decreases in total client assets and find similar results.
[14] Our results are also consistent with prior research that finds cost savings (i.e., "low-balling") for clients that switch auditors (e.g., DeAngelo 1981; Dye 1991). While these studies suggest a threat to auditor independence, evidence of lower audit quality is sparse (DeFond and Zhang 2014). In contrast, our findings of lower costs resulting from economies of scale suggest audit quality would be preserved.

to non-Big-4 auditors could indicate Big-4 audit offices' ability to easily leverage other offices' industry expertise to achieve cost advantages from economies of scale quicker.

For each analysis, we also present the relation for both types of audit firms on the same graph to provide some context for comparison. These comparative analyses are presented for observations with up to seven clients, which allows us to "zoom in" on the converging cost curves while still capturing most observations. Upon comparing Big-4 and non-Big-4 auditors (Panel C), the level of average fees per assets is significantly higher for non-Big-4 auditors in the left side of the distribution. At the right side of the distribution, the average fee per million dollars of assets audited by non-Big-4 auditors appears to converge to that of Big-4 auditors.

[Insert Figure 2 Here]

National-Industry Level

Next, we analyze economies of scale at the national-industry level. Our results, presented in Figure 3, provide strong support for economies of scale at the national-industry level. Notably, this result is in contrast to our mean difference results, which were not able to detect national-level economies of scale on average. This highlights the importance of using nonparametric analysis in detecting relations beyond average effects. For Big-4 firms (Panel A), the average cost curve appears to closely follow the theorized U-shape. Average fees per million dollars of assets are minimized around 20 clients in an industry in a given year and remain relatively flat up to 60 or 80 clients, at which point they begin to increase again. This could represent diseconomies of scale or simply noise in the underlying data due to the decreasing number of observations that identify the cost curve in the right tail of the distribution (there are 4,080 observations with fewer than 60 clients and only 194 observations with 60 or more clients).

Non-Big-4 auditors (Panel B) have far fewer clients per industry per year on average, but they also show evidence of national-industry economies of scale. For non-Big-4 firms, average fees per dollar of assets appear to be minimized around five clients in an industry-year. Interestingly, at the national-industry level, the average cost curve for non-Big-4 firms is minimized at a smaller number of clients than that of Big-4 firms, despite the two groups converging to a similar average fee. Unlike for Big-4 firms, the average cost curve remains flat for non-Big-4 firms and does not seem to increase at the right side of the distribution. This could simply be because even the largest non-Big-4 auditors in an industry-year audit far fewer clients than Big-4 auditors.

[Insert Figure 3 Here]

City Level

In our last set of economies of scale analyses, we analyze economies of scale within cities, regardless of industry. The results, presented in Figure 4, provide strong evidence of economies of scale at the city level for both Big-4 and non-Big-4 auditors. In both groups, we observe a sharp decline in the average fee per dollar of assets as the smallest auditors take on additional clients in a given city. For Big-4 auditors (Panel A), average fees reach their minimum around 10 clients in a city, after which point the average cost curve remains relatively flat. For non-Big-4 auditors, average fees reach their minimum around five clients. Unsurprisingly, the cost curves reach their minimums at a greater number of clients in the city-level specification than in the city-industry specification as city-industries comprise cities more broadly. Once again, the average cost curve for non-Big-4 firms reaches its minimum at a smaller number of clients than that of Big-4 firms, consistent with Big-4 firms auditing more diverse and complex clients that require more tailored (i.e., less standardized) procedures. Panel C compares the two classes of auditors and reveals that average fee per dollar of assets is higher for non-Big-4 auditors but that the two curves again converge around seven clients, although there is more separation than in Figure 6 at the city-industry level.

[Insert Figure 4 Here]

Discussion

Taken as a whole, our findings suggest that economies of scale arise not only at the city-industry level, but also at the national-industry level and the overall city level. This implies that the cost savings that give rise to economies of scale are not only industry-specific and that knowledge held by both the audit office and the firm as a whole contributes to economies of scale. Bills, Cunningham, and Myers (2015) argue that non-Big-4 audit firms benefit from membership in associations that allow access to other member firms' expertise, joint trainings, and shared resources. This is one possible explanation for our finding that economies of scale are present in the audit market for non-Big-4 firms and may also explain why fees per dollar of assets audited are significantly higher for audit firms that audit only one or two publicly traded clients.

In addition, our results provide evidence on the size at which average cost is minimized. While there is some evidence of diseconomies of scale in the national-industry analysis, we do not generally observe increasing average costs at the right side of the curve. *A priori*, the number of clients audited by a firm is an endogenous choice. While in some cases a firm may attempt to take on more clients than is optimal, it seems unlikely this practice is prevalent. It is important to note that while we document the presence of economies of scale, our results do not suggest that audit firms/offices necessarily realize scale-based savings. The audit market is highly competitive. Our descriptive statistics suggest that the majority of small audit firms in our sample have only one or two publicly traded clients in a city and/or industry. It may not be realistic for these firms to take on the number of public clients required to realize economies of scale. Thus, our results should be interpreted as evidence of economies in scale in the audit market and not indicative that any particular audit firm is likely to gain economies of scale.

V. RE-EXAMINING AUDIT FEES, SPECIALISTS, AND SCALE

Our findings in Section IV support the theorized nonlinear economies of scale and the use of nonparametric analysis. We find evidence of economies of scale at the city-industry, national-industry, and city levels. However, prior research primarily focuses on city-industry economies of scale and has not documented economies of scale at the broader national-industry or city levels. To better reconcile our findings with prior literature, we re-examine the results in Fung et al. (2012) (FGK). FGK is one of the few studies to examine economies of scale at the city-industry level in the US audit market. Their analyses differentiate between specialist and non-specialist auditors and measure scale using a percentile rank of market share in a city-industry among a sample of Big-4 auditors between 2000 and 2007.[15] FGK concludes that economies of scale are primarily concentrated in city-industry specialist auditors. To ensure our specifications are comparable to theirs, we begin by estimating their model (1):

$$\begin{aligned} LAF = \alpha &+ \beta_1 Spec_{it} + \beta_2 Scale_{it} + \beta_3 LTA_{it} + \beta_4 LSEG_{it} + \beta_4 CATA_{it} + \beta_5 Quick_{it} + \beta_6 DE_{it} \\ &+ \beta_7 ROI_{it} + \beta_8 Foreign_{it} + \beta_9 Opinion_{it} + \beta_{10} YE_{it} + \beta_{11} Loss_{it} + \beta_{12} AAClients_{it} \\ &+ \beta_{13} Citysize_{it} + Fixed\ Effects + \varepsilon \end{aligned} \tag{1}$$

FGK's variables of interest are *Spec* and *Scale* (and later, the interaction between *Spec* and *Scale*). *Spec* indicates industry specialization based on the audit office with the greatest total fees in the city-industry, and *Scale* is the decile-ranked audit office size based on number of clients in

[15] FGK examine several other cross-sections not mentioned here for brevity.

the city-industry. All variables are defined in Appendix A. We begin with our sample of 90,849 company-year observations and remove observations with missing data needed to calculate the control variables. We also remove all non-Big-4 observations, consistent with FGK. This results in a final sample of 26,765 company-year observations from 2002-2018. We divide the sample into two time periods; 2002-2007 and 2008-2018. We use the initial time period to replicate FGK and the additional time period to build upon their results. We report descriptive statistics of all variables used in Table 3. All variables are consistent with those reported by FGK.

[Insert Table 3 Here]

In Table 4, we report the results that correspond to FGK's Table 5 and 6 (post-SOX columns). Consistent with FGK, we find industry specialists earn a premium. *Scale* reduces this premium, though *Scale* alone is not significantly related to audit fees when the interaction with *Spec* is included. Our point estimates for *Spec*, *Scale*, and *Spec*Scale* are consistent with FGK in magnitude, direction, and statistical significance. Control variables also load similarly to the FGK estimates, but there is slightly more variation. Despite minor differences, it appears our baseline specification allows for direct comparison to prior literature, and we use this model to delve deeper into economies of scale in the audit market.

[Insert Table 4 Here]

For consistency with prior literature, the sample in the above tests ends prior to the Great Recession. However, the audit market has changed over time (Audit Analytics 2020, 2021). In Columns (3) and (4), we extend our sample beyond 2007 to examine how the audit marketplace has evolved between 2008 and 2018. In these columns, the coefficients on *Spec* and *Spec*Scale* are still statistically significant with the expected sign, but the point estimates are generally reduced. The coefficient on *Scale* increases in absolute magnitude and becomes statistically significant from Column (2) to Column (4), suggesting that economies of scale are no longer limited to specialists. In addition, the coefficient on *Spec*Scale* is less negative in the extended sample than the original FGK sample, possibly indicating that specialist auditors realize lesser cost savings or pass on a smaller portion of their savings to clients than in the early portion of the sample period. Thus, it appears that over time, the dynamics of audit pricing among Big-4 firms have changed. As one would expect, using the full sample period in Columns (5) and (6), the point estimates are a blend of the two time periods. Consistent with our nonparametric analysis, these results suggest that economies of scale are present beyond city-industry specialists.

After replicating and confirming FGK's results, we consider whether it is possible to identify broader economies of scale in the FGK setting by adapting the model to better capture the theorized nonlinear relation between scale and audit fees. To do this, we extend our previous analysis by adding nonlinear terms to the OLS regression models, as suggested by Shipman, Swanquist, and Whited (2017). We begin by adding nonlinear terms to FGK's OLS specifications to examine how they may influence our inferences. It is important to note that we are not re-examining these specifications to re-test their hypothesis. We do not disagree with FGK's finding that economies of scale are present among city-industry specialist auditors. Rather, we suggest that this linear specification may not completely capture the theorized nonlinear economies of scale and thus additional analyses are warranted.

In Table 5, we include squared and cubic terms for the variables of interest (Columns (1) and (4)), control variables (Columns (2) and (5)), or all variables (Columns (3) and (6)) for both the original 2002- 2007 sample period (Columns 1-3) and the extended 2002- 2018 sample period (Columns 4-6). For brevity we only present the variables of interest. Because specialist is an indicator, we cannot include higher order terms. When the higher-order terms are included, *Scale*, *Scale*^{2} and *Scale*^{3} are significant, consistent with economies of scale (Figure 1) at the city level. Moreover, *Spec***Scale*, *Spec***Scale*^{2} and *Spec***Scale*^{3} (Columns (1), (3), (4), and (6)), are insignificant, suggesting that economies of scale at the city-industry level are no longer detected.[16] This finding indicates that the ability to identify economies of scale at the city or city-industry level is highly dependent on the model used. Further, as noted previously, OLS may not be the appropriate tool in this context. This all points to possible functional form misspecification (Shipman et al. 2017) and further supports the use of nonparametric analysis.

[Insert Table 5 Here]

Because of concerns over functional form misspecification, we more closely examine the different dimensions of audit office size in the OLS model. Specifically, the variables measuring audit office scale, industry specialization, and city size all capture audit office size to some extent. Whited et al. (2022) note econometric issues when several proxies capture a similar construct, which they refer to as "same construct controls." Audit office scale is a percentile rank of city-industry number of clients, industry specialists are defined as city-industry leaders based on audit fees, and city size measures the audit office (i.e., not overall city) size based on audit fees. Given that all three of these measures are based on the individual audit office, they each serve as proxies for audit office size (i.e., same construct controls). This suggests that the interaction between audit office scale and industry specialization may be capturing nonlinearity in the relation between scale and audit fees rather than two distinct constructs.

[16] In Columns (2), (3), (5), and (6), squared and cubic terms for a number of control variables are statistically significant (untabulated). Additionally, when we mean-center the variables of interest before squaring and cubing, we still find significant nonlinearity in the relation between *Scale* and audit fees, though the specific coefficients are altered somewhat.

We empirically investigate whether the findings in Table 4 may be unintentionally capturing nonlinearity by estimating a parsimonious version of equation (2), varying the inclusion of *Spec*, *Scale*, and *Citysize*. The results of these specifications are presented in Table 6. When we omit any two of these three variables, the coefficient on the third is significantly positive. When we include both *Spec* and *Scale* but omit *Citysize*, we find no significant correlation between *Scale* and audit fees. This suggests that the negative relation between audit fees and *Scale* is driven by the inclusion of *Citysize*. Recall that *Spec* is calculated by ranking audit firms' total fees in each city-industry, *Scale* is calculated by ranking audit firms' number of clients in each city-industry, and *Citysize* is the logarithm of each audit firm's total fees in each city.

The results of our parsimonious models can be interpreted to suggest that audit firms that are specialists in a given city-industry do not pass on scale discounts to their clients unless they have a larger presence in the city overall, irrespective of industry. This result is consistent with our findings in Figure 4 that document city-level economies of scale irrespective of industry. This result also highlights the need to consider the impact of nonlinearity, as *Spec, Scale*, and *Citysize* each capture different facets of audit office size. Thus, their interactions may function as indirect proxies for higher-order size terms.

[Insert Table 6 Here]

VI. Additional Analyses

We perform a number of supplemental tests and analyses to provide additional insight into economies of scale in the audit market. First, in our main economies of scale analyses (Figures 2-4), we partition the sample into Big-4 and non-Big-4. While this partition is common in the literature, some studies further separate "second-tier" auditors from the non-Big-4 sample (e.g., Boone et al. 2010). We define second-tier audit firms as BDO and Grant Thornton and replicate our nonparametric analysis (untabulated). Inferences in the non-Big-4 audit sample remain largely unchanged. In both the city and city-industry specifications, second-tier audit firms demonstrate the classical downward-sloping nonlinear cost curve relative to their number of clients. In the national-industry specification, significantly more nonlinearity is present. We attribute this result to a large restructuring at BDO during our sample period.

Next, we perform a number of untabulated analyses to ensure our findings are not an artifact of research design choices. In our main analyses, we rely on Stata's default settings. Our inferences are unchanged when we re-estimate Figures 2-4 by altering the bandwidth and kernel weighting function and when we remove the financial services industry.

VII. CONCLUSION

In this paper, we revisit economies of scale in the audit market to reconcile economic theory with the results of prior accounting studies. Using nonparametric techniques and a novel measure of average cost, we detect economies of scale at the city, city-industry, and national-industry levels, typically arising with fewer than 10 public clients. We reexamine Fung et al. (2012) and find evidence that same construct controls bias the models' coefficient estimates. After extending tests from FGK to capture nonlinearity, we find audit firms that are specialists in a given city-industry do not pass on scale discounts to clients unless they have a larger presence in the city overall, irrespective of industry. Even without industry specialization, large audit firms are able to pass on some economies of scale savings to clients. Our evidence suggests the inability of prior research to document scale discounts outside of Big-4 industry specialists is attributable to research design choices rather than the absence of scale economies.

Our paper contributes to both the methodological literature in accounting and the literature on economies of scale in the audit market. While economies of scale are often assumed to exist, prior studies have been unable to document their existence beyond limited settings. We use a measure that captures average cost and thus more closely conforms to the theoretical model to document economies of scale in broad settings across the audit market. Our analyses raise questions about the functional form and broader empirical approach employed in the existing audit literature. By using nonparametric techniques, we are able to examine average costs across the distribution of audit firm/office sizes and thus document the firm/office size at which economies of scale are realized. In doing so, our analysis generates normative suggestions on the level of output at which an audit firm may expect average costs to be minimized and informs regulators and audit firms on optimal market structure.

REFERENCES

Audit Analytics. 2020. *Eighteen Year Review of Audit Fee & Non-Audit Fee Trends*. Available at: https://www.auditanalytics.com/doc/AuditAnalytics_AuditFeeReport_Nov2020.pdf

Audit Analytics. 2021. *Who Audits Public Companies – 2021 Edition*. Available at: https://blog.auditanalytics.com/who-audits-public-companies-2021-edition/

Basioudis, I. G., and J. R. Francis. 2007. Big-4 audit fee premiums for national and office-level industry leadership in the United Kingdom. *Auditing: A Journal of Practice & Theory, 26*(2), 143-166.

Baumol, W. J., J.C. Panzar, and R. D. Willig. 1986. On the theory of perfectly-contestable markets. In J. E. Stiglitz and G.F. Mathewson (Eds.), *New developments in the analysis of market structure*, 339-370. London: Palgrave Macmillan.

Beatty, R. P. 1993. The economic determinants of auditor compensation in the initial public offerings market. *Journal of Accounting Research, 31*(2), 294-302.

Bills, K. L., D. C. Jeter, and S. E. Stein. 2015. Auditor industry specialization and evidence of cost efficiencies in homogeneous industries. *The Accounting Review* 90(5): 1721-1754.

Burgstahler, D. and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics, 24*(1), 99–126.

Cahan, S. F., D. C. Jeter, and V. Naiker. 2011. Are all industry specialist auditors the same? *Auditing: A Journal of Practice Theory* 30(4): 191-222.

Cameron, A. and P. Trivedi. 2005. *Microeconometrics: Methods and Applications.* New York: Cambridge University Press.

Caves, D., L. Christensen, and M. Tretheway. 1984. Economies of density versus economies of scale: why trunk and local service airline costs differ. *RAND Journal of Economics*, *15*(4), 471-489.

Chan, P., M. Ezzamel, and D. Gwilliam. 1993. Determinants of audit fees for quoted UK companies. *Journal of Business Finance and Accounting* 20: 765-786.

Chaney, P. K., D. C. Jeter, and L. Shivakumar. 2004. Self-selection of auditors and audit pricing in private firms. *The Accounting Review, 79*(1), 51-72.

Choi, J-H., C. F. Kim, J-B. Kim, and Y. Zang. 2010. Audit office size, audit quality, and audit pricing. *Auditing: A Journal of Practice & Theory* 29(1): 73-97.

Clatworthy, M. A., G. H. Makepeace, and M. J. Peel. 2009. Selection bias and the Big Four premium: New evidence using Heckman and matching models. *Accounting and Business Research, 39*(2), 139-166.

Craswell, A., J. Francis, and S. Taylor. 1995. Auditor brand name reputations and industry specializations. *Journal of Accounting and Economics* 20: 297-322.

DeAngelo, L. 1981. Auditor independence, "low-balling" and disclosure regulation. *Journal of Accounting and Economics* 3: 113-127.

DeFond, M. L, D. H. Erkens, and J. Zhang. 2016. Do client characteristics really drive the Big N audit quality effect? New evidence from propensity score matching. *Management Science*, *63*(11), 3628-3649.

DeFond, M. L., J. R. Francis, and T. J. Wong. 2000. Auditor industry specialization and market segmentation: Evidence from Hong Kong. *Auditing: A Journal of Practice & Theory*, *19*(1), 49-66.

DeFond, M. and J. Zhang. 2014. A review of archival auditing research. *Journal of Accounting and Economics, 58*(2-3), 275–326.

Dye, R. 1991. Informationally motivated auditor replacement. *Journal of Accounting and Economics* 14: 347-374.

Fan, J. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, *87*(420), 998-1004.

Fan, J. and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Francis, J. R. 1984. The effect of audit firm size on audit prices: A study of the Australian market. *Journal of Accounting and Economics* 6(2): 133-151.

Francis, J. R., M. N. Mehta, and W. Zhao. 2017. Audit office reputation shocks from gains and losses of major industry clients. *Contemporary Accounting Research* 34(4): 1922-1974.

Francis, J. R., K. Reichelt, and D. Wang. 2005. The pricing of national and city-specific reputations for industry expertise in the U.S. audit market. *The Accounting Review, 80*(1), 113–136.

Francis, J. R., and D. Simon. 1987. A test of audit pricing in the small-client segment of the U.S. audit market. *The Accounting Review* 62: 145-1557.

Francis, J. R., and D. J. Stokes. 1986. Audit prices, product differentiation, and scale economies: Further evidence from the Australian market. *Journal of Accounting Research, 24*(2), 383-393.

Freeman R. N., and S. Y. Tse. 1992. A nonlinear model of security price responses to unexpected earnings. *Journal of Accounting Research, 30*(2), 185-209.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically, *The Review of Financial Studies, 33*(5) 2326-2377.

Fung, S. Y. K., F. A. Gul, and J. Krishnan. 2012. City-level auditor industry specialization, economies of scale, and audit pricing. *The Accounting Review, 87*(4), 1281–1307.

Gong, Q., O. Zhen, Y. Lin, and L. Wu. 2016. On the benefits of audit market consolidation: Evidence from merged audit firms. *The Accounting Review* 91(2): 463-488.

Gow, I. D., D. F. Larcker, and P. C. Reiss. 2016. Causal inference in accounting research. *Journal of Accounting Research, 54*(2), 477–523.

Härdle, W. and O. Linton. 1994. *Handbook of Econometrics* (4th ed.). Amsterdam: Elsevier, (Chapter 38).

Hay, D. 2013. Further evidence from meta-analysis of audit fee research. *International Journal of Auditing, 17*(2), 162-176.

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

Hogan, C. E., and M. S. Wilkins. 2008. Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies? *Contemporary Accounting Research, 25*(1), 219–242.

Jones, M. C., J. S. S. Marron, and S. J. J. Sheather. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association, 91*(433), 401–407.

Kennedy, P. 2008. *A guide to econometrics*. John Wiley & Sons.

Lawrence, A., M. Minutti-Meza, and P. Zhang. 2011. Can Big-4 versus non-Big-4 differences in audit-quality proxies be attributed to client characteristics? *The Accounting Review*, *86*(1), 259–286.

Lennox C. S., J. R. Francis, and Z. Wang. 2012. Selection models in accounting research. *The Accounting Review, 86*(2), 589-616.

Mayhew, B. W., and M. S. Wilkins. 2003. Audit firm industry specialization as a differentiation strategy: Evidence from fees charged to firms going public. *Auditing: A Journal of Practice & Theory, 22*(2), 33-52.

McMeeking, K. P., K. V. Peasnell, and P. F. Pope. 2006. The determinants of the UK big firm premium. *Accounting and Business Research, 36*(3), 207-231.

Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability & Its Applications*, *9*(1), 141-142.

Opsomer J. D., and F. J. Breidt. 2011). Nonparametric regression using kernel and spline methods. In Lovric M. (Ed,), *International Encyclopedia of Statistical Science*. Berlin: Springer.

Panzar, J. C., and R. D. Willig. 1981. Economies of scope. The American Economic Review, 71(2), 268-272.

Pearson, T. and G. Trompeter. 1994. Competition in the market for audit services: The effect of supplier concentration on audit fees. *Contemporary Account. Research, 11*(1), 115-135.

Reichelt, K. J., and D. Wang. 2010. National and office-specific measures of auditor industry expertise and effects on audit quality. *Journal of Accounting Research* 48(3): 647-686.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 832-837.

Schelleman, C., and W. R. Knechel. 2010. Short-term accruals and the pricing and production of audit services. *Auditing: A Journal of Practice & Theory, 29*(1), 221-250.

Shipman, J. E., Q. T. Swanquist, and R. L. Whited. 2017. Propensity score matching in accounting research. *The Accounting Review, 92*(1), 213–244.

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

Stigler, G. J. 1958. The economies of scale. *The Journal of Law & Economics*, 1, 54–71.

Watson, G. S. 1964. Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A, 26(4), 359-372.

Whited, R. L., Q. T. Swanquist, J. E. Shipman, and J. R. Moon. 2022. Out of Control: The (Over) Use of Controls in Accounting Research. *The Accounting Review* 97 (3): 395–413.

Zerni, M. 2012. Audit partner specialization and audit fees: Some evidence from Sweden. *Contemporary Accounting Research* 29(1): 312-340.

APPENDIX A

Variable Definitions

Variable Name	Definition
Fees per Asset	Total audit fees divided by total assets (in millions) of clients of an audit firm in either the city, industry, or city-industry
NClients	Number of unique clients for an audit firm at either the city-industry, national-industry, or city level
LAF	Log of audit fees (*audit fees*)
Spec	Indicator variable equal to 1 if the audit firm has the largest market share in the city-industry and 0 otherwise
Scale	Percentile rank of the number of audit clients (*NClients*) in a city-industry for each Big-N audit firm
LTA	Log of total assets (*AT*)
LSEG	Log of the number of unique business segments
CATA	Ratio of current assets to total assets (*ACT*/AT*)*
Quick	Ratio of current assets (excluding inventories) to current liabilities ((*ACT-INV*)/*LCT*)
DE	Ratio of long-term debt to total assets (*DLTT*/*AT*)
ROI	Ratio of earnings before interest and taxes to total assets (*EBIT*/*AT*)
Foreign	Ratio of foreign sales to total sales
Opinion	Indicator variable equal to 1 if there is a going concern opinion and 0 otherwise
YE	Indicator variable equal to 1 for non-December fiscal year ends and 0 otherwise
Loss	Indicator variable equal to 1 for negative net income and 0 otherwise
AAclients	Indicator variable equal to 1 if there was an auditor change and the audit firm in year *t-1* was Arthur Andersen and zero otherwise
Citysize	Natural log of aggregate audit fees for all firms audited by the company's auditor for each city

APPENDIX B

Econometric Detail

This Appendix provides a more detailed econometric discussion of nonparametric analyses. We provide the background, in terms common to accounting researchers, necessary to utilize these techniques. We follow Cameron and Trivedi (2005) Chapter 9 to obtain the estimator of a histogram through the statistical properties of the estimate (Section 9.3).[17] For a sample of N observations $\{x_i, i = 1, \dots, N\}$, consider the continuous variable x_0, which is evaluated at point x and has a distribution function $f(x)$. Furthermore, the variable h, bin width, will help determine the range of points that will be included in a bin. Applied to the Burgstahler and Dichev (1997) example, x represents earnings, and h represents the distance to the midpoint of the earnings bin (.0025/2). The histogram is estimated with the following equation:

$$\hat{f}_{\text{Hist}}(x_0) = \frac{1}{N} \sum_{i=1}^{N} \frac{1(x_0 - h < x_i < x_0 + h)}{2h} \qquad (A1)$$

The indicator function is equal to one if the observation occurs within the range $x_0 \pm h$. This equation yields a step function (based on each bin) that weights all observations within a given bin equally. We can now extend equation (A1) to study a case in which we weight each point within a bin or bandwidth (bandwidth in nonparametric analysis is analogous to bin width in a histogram). This approach weights points using a kernel weighting function based on the premise that points closer to x are likely more representative and should be considered more than points further from x. In terms of Burgstahler and Dichev (1997), this would be equivalent to suggesting that companies with earnings closest to zero are more similar to one another than the companies toward either tail of this bin. It is important to note that the final kernel destiny function fits a smooth

[17] We generally refer to asymptotic properties, unless otherwise noted.

estimate based on the estimation at each point, x_i.[18] This analysis also relies on several assumptions about the kernel weighting function, *K(·)*. These assumptions include symmetry around zero, continuity, and integration to one. The kernel density estimate can then be written as:

$$\hat{f}(x_0) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{1(x_i - x_0)}{h}\right) \qquad (A2)$$

While this function looks similar to that of a histogram, equation (A2) uses a weighting function, *K(·)*, instead of summing a series of indicators over a particular range. Given the equation for kernel density estimates, there are two primary research design choices: bandwidth, *h*, and the weighting function, *K(·)*. The bandwidth is the distance around data points that will be evaluated by the weighting function and is analogous to bin width in a histogram. The weighting function specifies how much weight to assign to particular observations within the bandwidth. In general, the choice of bandwidth is more important than the choice of weighting function, because many weighting functions have similar statistical properties.

While understanding the distribution of a single variable is useful in many settings, we next turn our attention to the case of univariate regression.[19] Freeman and Tse (1992) document that the relation between securities prices and unexpected earnings is nonlinear. Nonparametric analysis would allow us to reexamine this association without the assumption, as in Freeman and Tse (1992), that the functional relation between these two variables follows an inverse tangent distribution. We could estimate the following regression equation, where *y* is the unexpected return and *x* is unexpected earnings:

---

[18] While most statistical packages use the same evaluation criteria for bandwidths across a specific number of points along the entire distribution, the analysis can theoretically be performed pointwise.

[19] For tractability, we limit discussion to constant and local linear regression. In this section we continue to assume that both x and y have positive econometric properties (that is, that they are IID, continuous, twice differentiable, and have finite variances).

$$y_i = m(x_i) + \varepsilon_i \qquad (A3)$$

Following Cameron and Trivedi (2005), we describe the estimation procedure and document the general case, in which we estimate the function $m(\cdot)$ by taking the average values of y_i for points that are within h of each observation x:[20]

$$\widehat{m}(x_0) \equiv \frac{\sum_{i=1}^{N} 1\left(\left|\frac{x_i - x_0}{h}\right| < 1\right) y_i}{\sum_{i=1}^{N} 1\left(\left|\frac{x_i - x_0}{h}\right| < 1\right)} \qquad (A4)$$

Similar to a histogram, this equation assigns equal weight to all observations that fall within the bandwidth h. If we wish to vary the weights placed on observations that are closer to x_0, we need a weighting function. In this case we use a kernel weighting function, as above, and obtain the following estimate of $m(\cdot)$:

$$\widehat{m}(x_0) \equiv \frac{\frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x_i - x_0}{h}\right) y_i}{\frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x_i - x_0}{h}\right)} \qquad (A5)$$

This regression technique estimates a constant for each value of x and then plots a curve via interpolation (Opsomer and Breidt (2011)). We can extend this technique beyond a constant to estimate a weighted local linear model (Fan (1992)).[21] To show this extension, we turn our attention to the estimation of the following functional form of $m(\cdot)$:

$$m(x) = \alpha_0 + \beta_0 (x - x_0) \qquad (A6)$$

To generate a local linear regression, we set α and β to minimize:

$$\sum_{i=1}^{N} K\left(\frac{x_i - x_0}{h}\right) (y_i - \alpha_0 - \beta_0 (x_i - x_0))^2 \qquad (A7)$$

[20] This estimator can be derived by integrating the joint density function of x and y. See Härdle and Linton (1994).
[21] Higher order polynomials can be estimated in a similar fashion. See Härdle and Linton (1994) equation (21) or Cameron and Trivedi (2005) equation (9.31). See Fan and Gijbels (1996) for properties of these higher order polynomial estimators.

As in the density estimation analysis, the researcher chooses the bandwidth, h, and the weighting function, $K(\cdot)$. Equation (A7) can be estimated using weighted least squares at each point x_0. This yields a smooth linear estimate around each observation x. Equation (A7) produces the locally weighted linear estimate of $\widehat{m}(x)$. It is important to note that kernel density regression does not handle end points as well as locally weighted linear regressions (Fan (1992)).

Several other econometric techniques warrant discussion. The first of these is a technique called k-nearest neighbor estimation (often abbreviated as k-NN or KNN). Instead of focusing on the area around a point, X_0, based on the bandwidth, h, this technique uses a weighted average of the values of the k closest observations to the point. Another nonparametric technique is the use of splines. Opsomer and Breidt (2011) describe splines as "an alternative approach… to represent the fit as a piecewise polynomial, with the pieces connecting at points called knots" (p. 976). This methodology has recently been used in the capital markets literature to examine cross-sectional attributes that explain equity returns in Freyberger, Neuhierl, and Weber (2020).[22] A full description of these estimators is beyond the scope of this paper.

Choice of Weighting function and Bandwidth

We next demonstrate how research design choices of weighting function and bandwidth affect estimation. While there are numerous options for kernel weighting functions, for brevity, we discuss two: Epanechnikov and Guassian. Under certain assumptions, the Epanechnikov kernel is the most efficient and, accordingly, Stata uses the Epanechnikov kernel as the default setting. The Gaussian kernel is also important because, unlike other kernels (e.g., Epanechnikov, among others), it gives some weight to observations that fall outside the bandwidth. This feature reduces

[22]Freyberger, Neuhierl, and Weber (2020) uses splines in conjunction with a group LASSO procedure.

estimation concerns when there are gaps in the data. As we move outside the bandwidth, however, these points receive very little weight.[23]

In Figure A1, we demonstrate how differences in bandwidth affect the nonparametric estimation. To limit the number of dimensions we alter at one time, we limit this analysis to Big-N, city-level-specialist auditors (N=12,717). In Panel A, we compare two weighting functions to plot log audit fees for specialists. The empirical distributions appear remarkably similar using the Epanechnikov (i.e., default) and Gaussian kernels. While differences in weighting function do impact the estimation process, these differences are generally small unless there are holes in the data. To highlight this, in Panel B, we plot *Scale* using the two density functions. *Scale* is based on a discrete number of audits, creating large jumps in the percentile rank when moving from one to two to three clients. There are significantly fewer gaps in the right tail. Using these data, we see that the Gaussian kernel provides a smoother empirical distribution around points of zero probability mass by giving weight to points outside the bandwidth. This may or not be desirable depending on the setting. If the data are distributed normally and gaps in the data result from a random sampling process, the Gaussian kernel fills these voids to produce a full empirical distribution. On the other hand, if there are gaps in the data by construction, as with *Scale*, this procedure may mask these gaps, making this tool less informative from a diagnostic standpoint.

[Insert Figure A1 Here]

Next, we consider the choice of bandwidth. To do so, we systematically vary the bandwidth while holding constant the Epanechnikov kernel weighting function to plot the distribution of log audit fees and *Scale* for specialist auditors. We select three different bandwidths: one that

[23] Consider the example of a random variable centered on zero and normally distributed. The probability density function is defined from -∞ to ∞, but as we move into the tails of the distribution, the probability approaches zero.

undersmooths, the default setting, and one that oversmooths. In Figure A2, Panel A, we plot the distribution of log audit fees. When we undersmooth by setting the bandwidth smaller than optimal, the figure appears jagged, and it is difficult to assess whether the irregular shape of the curve is due to a true underlying pattern or to random unmeaningful fluctuations in the data. This image is consistent with the observation from Jones et al. (1996) that an undersmoothed estimate "is too rough." Alternatively, when we set too large a bandwidth, we essentially "smooth away" all detail, leaving a relatively symmetric curve providing very little insight into the distribution. When the optimal (default) bandwidth is used, the curve is neither oversmoothed nor undersmoothed.[24]

In Figure A2, Panel B, the differences caused by choice of bandwidth are striking. Most notably, when we undersmooth the distribution of *Scale,* points of zero probability mass emerge. In this case, undersmoothing serves as a diagnostic tool that demonstrates a potential drawback of using percentiles when data is right-skewed. These areas of zero probability mass are not apparent when we oversmooth. We reiterate the advice of Cameron and Trivedi (2005), which suggests halving and doubling the default bandwidth and reporting all three for robustness. Collectively, these figures document that, even with gaps in the data, bandwidth is a more important research design choice than kernel weighting function.

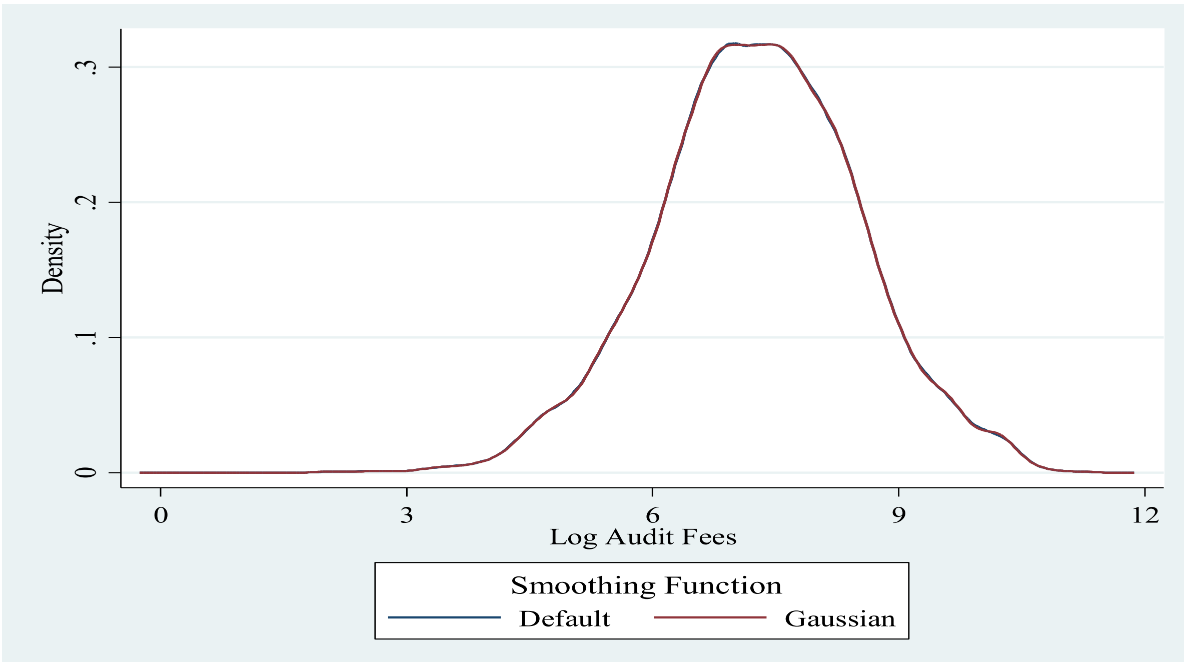
[Insert Figure A2 Here]

[24] As noted earlier, the optimal bandwidth minimizes the mean integrated squared error.

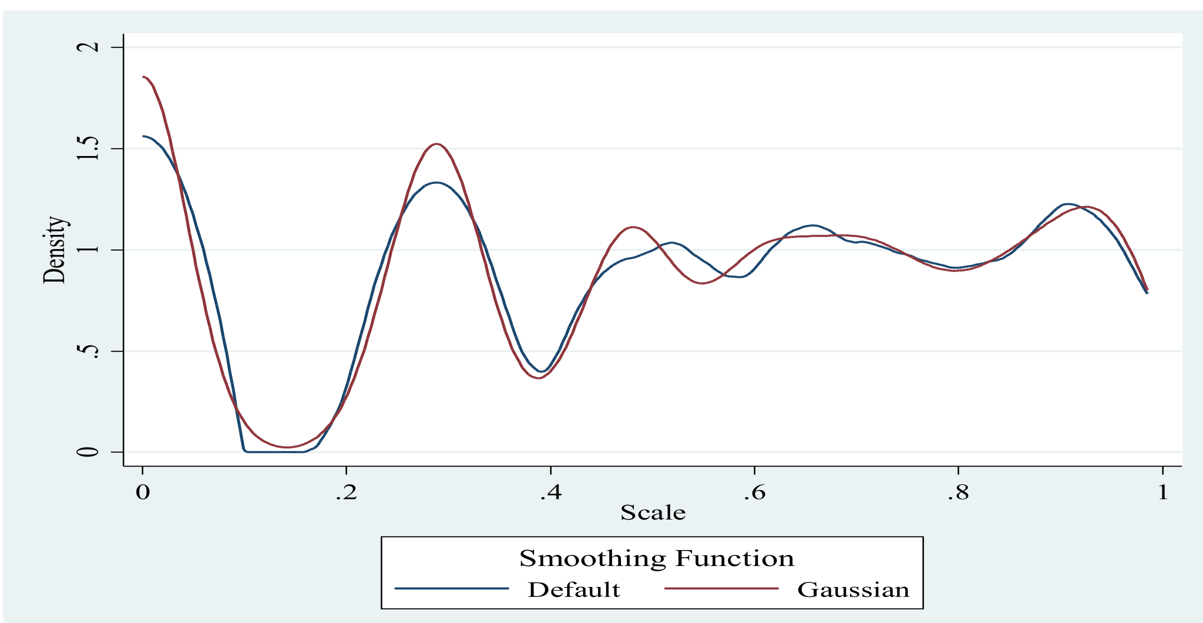
FIGURE A1

Choice of Weighting Function

Panel A: Audit Fees



Panel B: Industry Specialist Economies of Scale

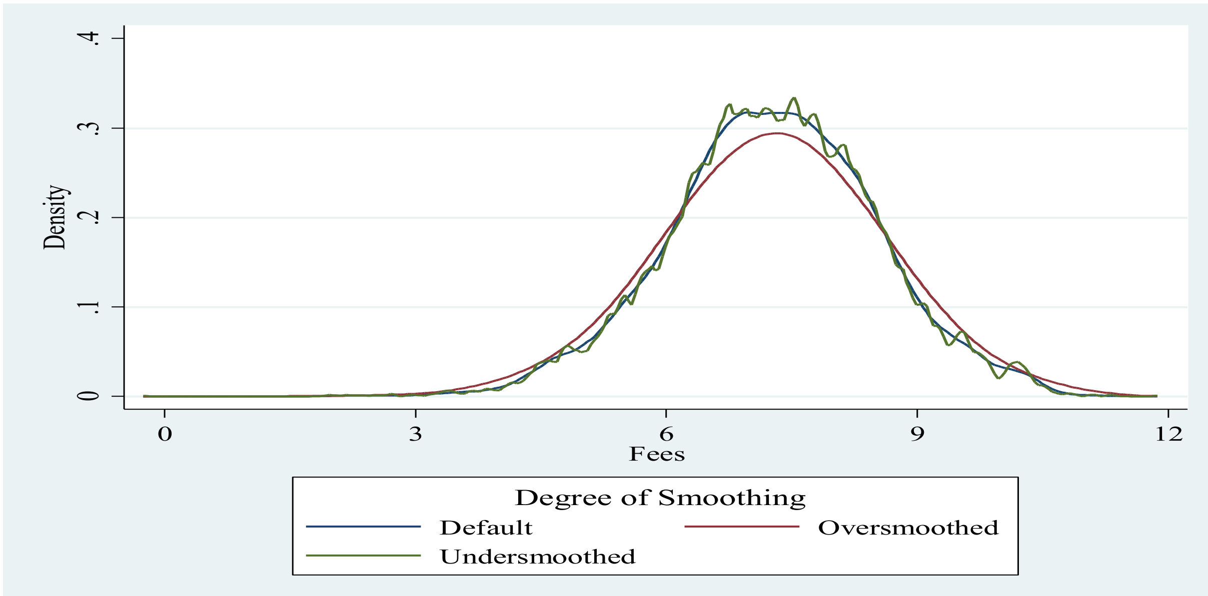


Note: These figures represent the choice of weighting within the kernel density estimation of log-transformed audit fees (Panel A) and *Scale* (Panel B). We limit the estimation of the empirical distributions to specialist Big-4 auditors. We consider two weighting options: the Epanechnikov (default) and Gaussian.

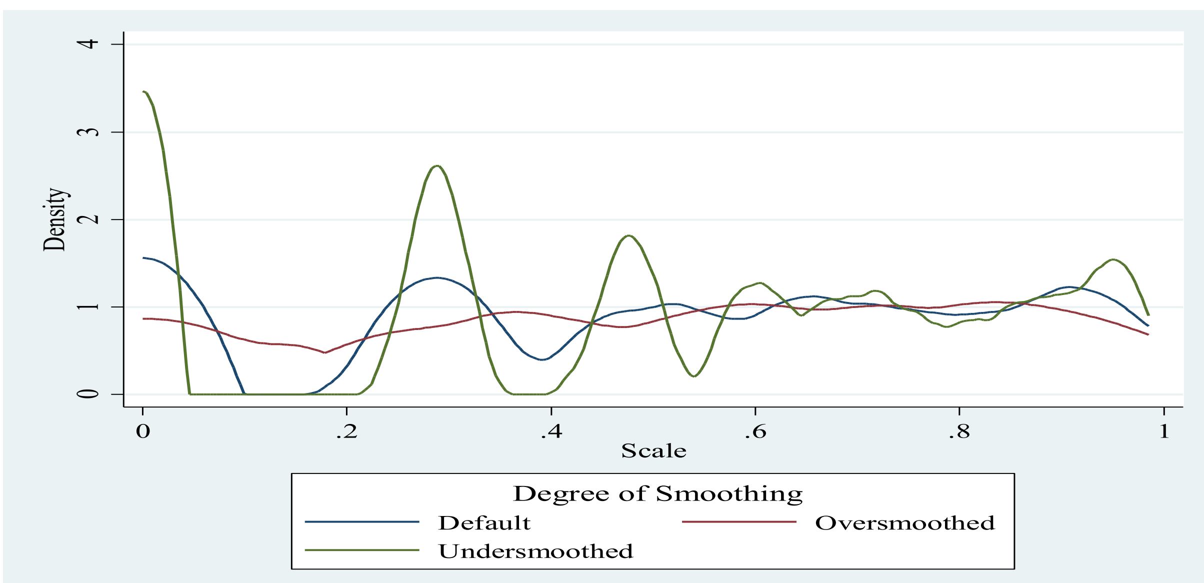
FIGURE A2

Choice of Bandwidth

Panel A: Audit Fees



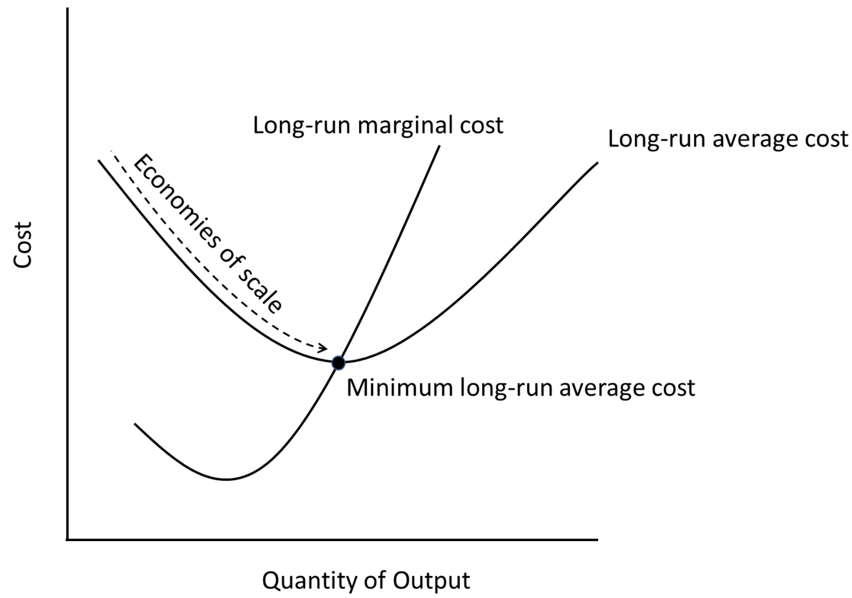
Panel B: Industry Specialist Economies of Scale



Note: These figures represent the choice of bandwidth within the kernel density estimation of log-transformed audit fees (Panel A) and *Scale* (Panel B). We limit the estimation of the empirical distributions to specialist Big-4 auditors. We select the bandwidth to default, over-, and under-smooth the distributions. The default bandwidth is the optimal bandwidth calculated and used in Stata based on minimizing the mean integrated squared error.

FIGURE 1

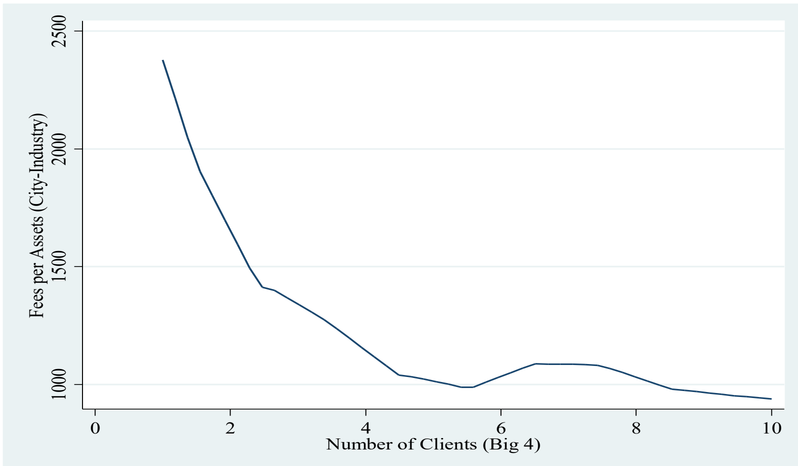
Economies of Scale



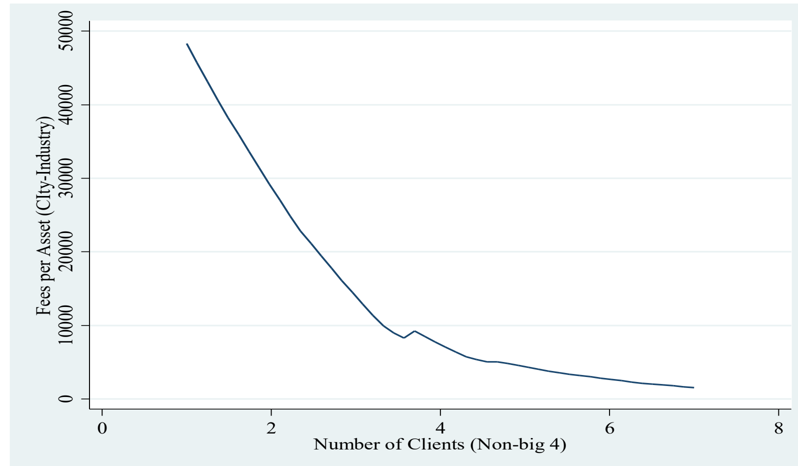
Note: This figure represents the theoretical relation between average cost and quantity of output, which forms the basis for economies of scale.

FIGURE 2
City-Industry Economies of Scale

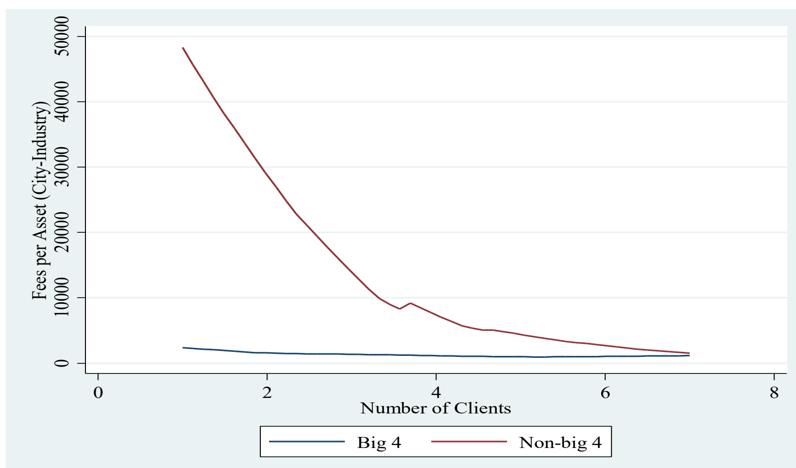
Panel A



Panel B



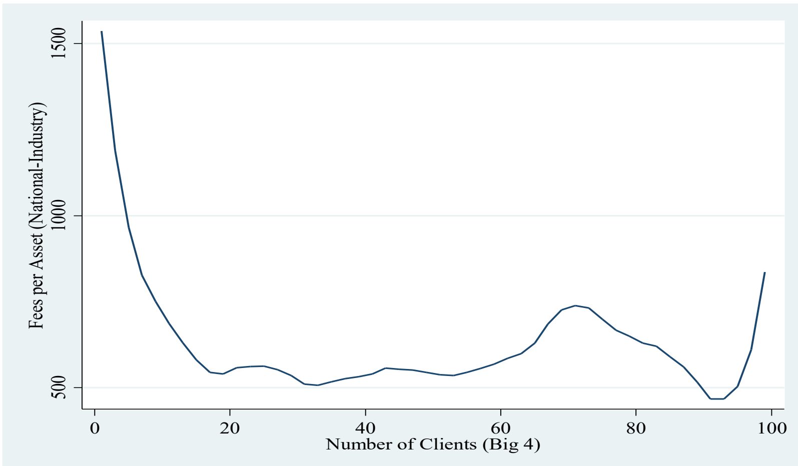
Panel C



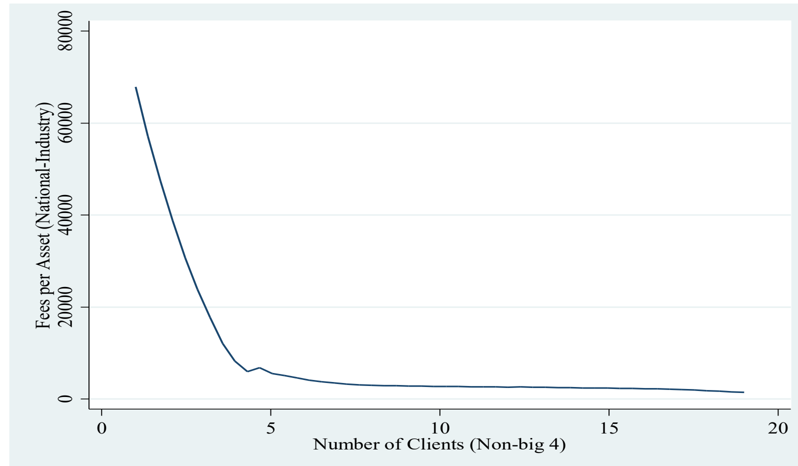
Note: These figures represent the nonparametric regression of audit fees per assets (aggregated at the audit firm-city-industry level) on the number of clients. We perform locally weighted linear regression with the default bandwidth and weighting function. In Panels A and B, we limit the distribution to the 99 percentile. In Panel C, we limit the sample to observations with seven or fewer clients.

FIGURE 3
National-Industry Economies of Scale

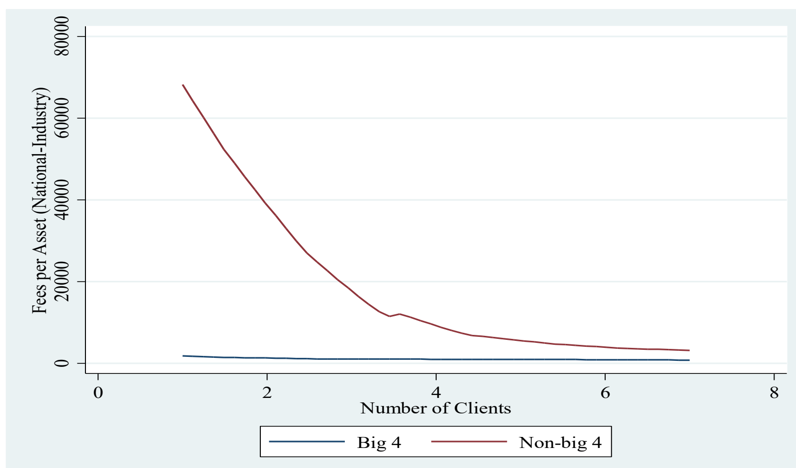
Panel A



Panel B



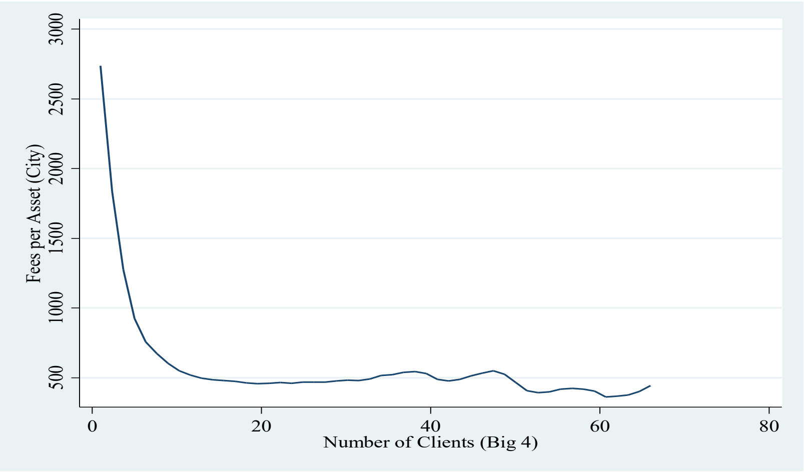
Panel C



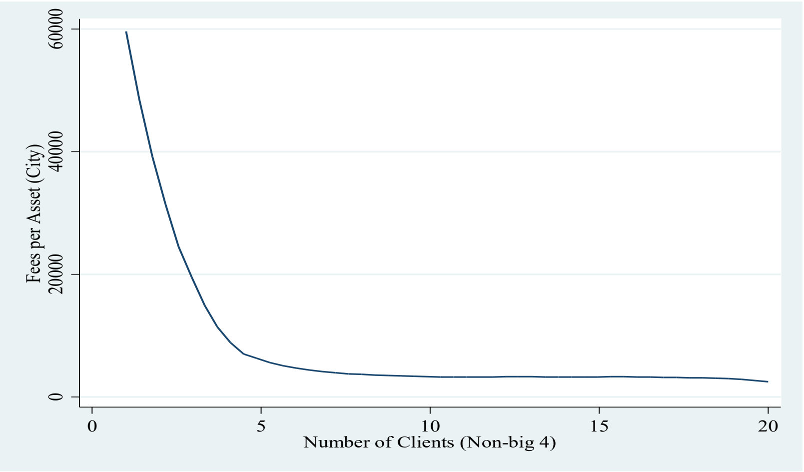
Note: These figures represent the nonparametric regression of audit fees per assets (aggregated at the national audit firm-industry level) on the number of clients. We perform locally weighted linear regression with the default bandwidth and weighting function. In Panels A and B, we limit the distribution to the 99 percentile. In Panel C, we limit the sample to observations with seven or fewer clients.

FIGURE 4
City Economies of Scale

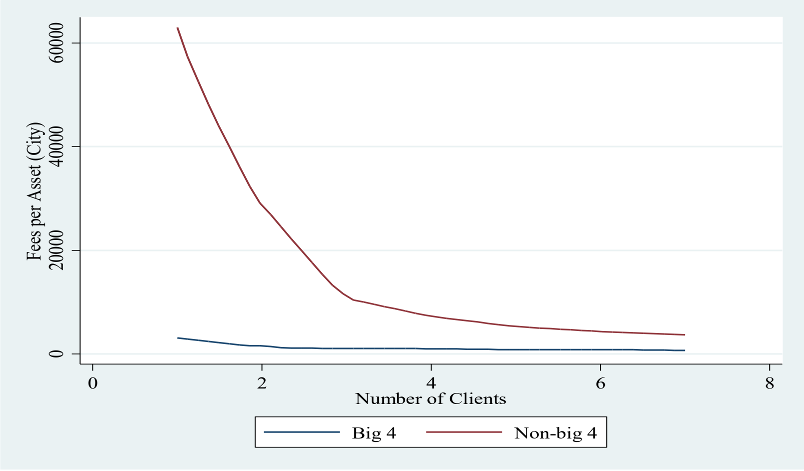
Panel A



Panel B



Panel C



Note: These figures represent the nonparametric regression of audit fees per assets (aggregated at the audit firm-city level) on the number of clients. We perform locally weighted linear regression with the default bandwidth and weighting function. In Panels A and B, we limit the distribution to the 99 percentile. In Panel C, we limit the sample to observations with seven or fewer clients.

TABLE 1

Descriptive Statistics: Economies of Scale Nonparametric Estimation

Panel A: Full Sample

	N	Mean	Std Dev	25th Pctl	50th Pctl	75th Pctl
City-Industry Level						
Fees per Asset	55,943	19,239	456,029	505	1,540	4,440
NClients	55,943	1.608	1.738	1	1	2
National-Industry Level						
Fees per Asset	21,305	43,904	736,790	850	2,837	8,709
NClients	21,305	4.252	10.760	1	1	3
City Level						
Fees per Asset	15,181	20,442	561,487	308	1,214	4,343
NClients	15,181	6.104	9.377	1	3	7

Panel B: Non-Big-4 vs. Big-4

	Non-Big-4			*Big-4*		
	N	Mean	Std Dev	N	Mean	Std Dev
City-Industry Level						
Fees per Asset	24,074	41,947	694,483	31,821	2,067	6,731
NClients	24,074	1.43	1.19	31,821	1.75	2.05
National-Industry Level						
Fees per Asset	17,003	54,777	824,399	4,279	934	1,387
NClients	17,003	2.04	3.75	4,279	13.06	20.58
City Level						
Fees per Asset	9,964	30,553	692,858	5,189	1,131	3,232
NClients	9,964	3.69	4.42	5,189	10.76	13.67

Note: All variables are defined in Appendix A.

TABLE 2

Changes in Audit Clients and Economies of Scale

Panel A: Net Increases to Public Client Base

	N	Mean *Fees per Asset (t)*	Mean *Fees per Asset (t-2)*	Mean Diff. Diff.	Mean Diff. t-stat
City-Industry Level					
Big-4	2,314	1,422	2,000	-578***	-3.37
Non-Big-4	1,573	17,902	28,160	-10,259	-0.78
National-Industry Level					
Big-4	818	741	736	4.48	0.20
Non-Big-4	1,623	17,551	29,028	-11,477	-0.89
City Level					
Big-4	1,026	738	975	-237***	-2.88
Non-Big-4	1,913	5,634	14,100	-8,466***	-3.28

Panel B: Net Decreases to Public Client Base

	N	Mean *Fees per Asset (t)*	Mean *Fees per Asset (t-2)*	Mean Diff. Diff.	Mean Diff. t-stat
City-Industry Level					
Big-4	3,264	2,152	1,430	721***	7.28
Non-Big-4	1,506	47,262	9,975	37,287***	2.61
National-Industry Level					
Big-4	1,620	910	821	89***	4.85
Non-Big-4	1,404	54,037	10,885	43,151***	2.79
City Level					
Big-4	1,956	920	712	209***	4.51
Non-Big-4	1,860	19,045	5,401	13,645*	1.66

Note: All variables are defined in Appendix A. *, **, and *** signify statistical significance at the 10%, 5%, and 1% significance level, respectively.

Table 3

Descriptive Statistics: Audit Fees, Specialists, and Scale

	N	**Mean**	**Std Dev**	**25th Pctl**	**50th Pctl**	**75th Pctl**
			Fiscal Years 2002-2007			
LAF	12,006	6.520	1.275	5.629	6.503	7.345
Spec	12,006	0.470	0.499	-	-	1.000
Scale	12,006	0.429	0.334	-	0.466	0.723
LTA	12,006	6.021	1.986	4.648	5.967	7.353
LSEG	12,006	1.165	0.889	0.693	0.693	2.079
CATA	12,006	0.506	0.257	0.299	0.509	0.713
Quick	12,006	2.627	6.655	0.952	1.509	2.773
DE	12,006	0.201	0.299	0.000	0.119	0.300
ROI	12,006	(0.028)	0.552	(0.023)	0.059	0.110
Foreign	12,006	0.198	0.338	-	0.006	0.349
Opinion	12,006	0.038	0.191	-	-	-
YE	12,006	0.292	0.455	-	-	1.000
Loss	12,006	0.366	0.482	-	-	1.000
AAclients	12,006	0.002	0.047	-	-	-
Citysize	12,006	19.843	1.514	19.067	20.248	20.887
			Fiscal Years 2002-2018			
LAF	26,765	6.980	1.243	6.205	7.005	7.783
Spec	26,765	0.475	0.499	-	-	1.000
Scale	26,765	0.427	0.336	-	0.466	0.727
LTA	26,765	6.550	2.040	5.170	6.545	7.932
LSEG	26,765	1.107	0.909	0.693	0.693	2.079
CATA	26,765	0.487	0.265	0.267	0.480	0.695
Quick	26,765	2.680	9.640	0.974	1.507	2.698
DE	26,765	0.220	0.311	0.001	0.157	0.331
ROI	26,765	(0.037)	0.898	(0.018)	0.060	0.109
Foreign	26,765	0.211	0.316	-	0.021	0.382
Opinion	26,765	0.038	0.190	-	-	-
YE	26,765	0.257	0.437	-	-	1.000
Loss	26,765	0.369	0.483	-	-	1.000
AAclients	26,765	0.001	0.032	-	-	-
Citysize	26,765	19.964	1.383	19.260	20.301	20.887

Note: Table 3 presents descriptive statistics for each our samples. All variables are defined in Appendix A.

TABLE 4
Audit Fees, Scale, and Specialization

DV=*LAF*	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	2002-2007	2002-2007	2008-2018	2008-2018	2002-2018	2002-2018
Spec	0.214***	0.397***	0.165***	0.276***	0.187***	0.333***
	(11.87)	(14.61)	(9.815)	(10.11)	(13.39)	(14.88)
Scale	-0.247***	-0.0644	-0.263***	-0.152***	-0.260***	-0.115***
	(-6.285)	(-1.437)	(-7.004)	(-3.637)	(-8.258)	(-3.296)
*Spec*Scale*		-0.412***		-0.250***		-0.328***
		(-8.308)		(-5.448)		(-8.534)
LTA	0.495***	0.490***	0.469***	0.465***	0.481***	0.476***
	(55.37)	(54.73)	(62.03)	(61.00)	(67.44)	(66.49)
LSEG	0.0728***	0.0729***	0.0611***	0.0617***	0.0690***	0.0695***
	(6.003)	(6.055)	(5.271)	(5.345)	(6.821)	(6.914)
CATA	0.321***	0.312***	0.327***	0.320***	0.326***	0.318***
	(5.070)	(4.965)	(5.554)	(5.440)	(6.583)	(6.424)
Quick	-0.00853*	-0.00839*	-0.00190	-0.00188	-0.00373	-0.00368
	(-1.729)	(-1.733)	(-0.872)	(-0.860)	(-1.412)	(-1.396)
DE	0.0325	0.0374	0.0901***	0.0906***	0.0652***	0.0675***
	(0.972)	(1.120)	(3.329)	(3.359)	(2.819)	(2.922)
ROI	-0.0915***	-0.0898***	-0.0316***	-0.0312***	-0.0401***	-0.0394***
	(-5.106)	(-5.135)	(-3.189)	(-3.184)	(-2.809)	(-2.808)
Foreign	0.236**	0.233**	0.393***	0.392***	0.306***	0.305***
	(2.158)	(2.176)	(10.54)	(10.56)	(3.893)	(3.915)
Opinion	0.422***	0.412***	0.230***	0.225***	0.325***	0.317***
	(9.299)	(9.169)	(5.900)	(5.764)	(10.23)	(10.03)
YE	-0.164***	-0.162***	-0.0150	-0.0151	-0.0891***	-0.0886***
	(-7.490)	(-7.472)	(-0.604)	(-0.612)	(-4.481)	(-4.495)
Loss	0.0986***	0.0941***	0.132***	0.129***	0.119***	0.115***
	(4.930)	(4.732)	(7.918)	(7.765)	(8.545)	(8.301)
AAclients	-0.155	-0.162			-0.200	-0.205
	(-1.099)	(-1.175)			(-1.399)	(-1.465)
Citysize	0.0845***	0.0875***	0.119***	0.120***	0.102***	0.104***
	(12.25)	(12.64)	(14.22)	(14.33)	(16.62)	(16.90)
Constant	1.007***	0.912***	1.384***	1.348***	0.703***	0.637***
	(7.251)	(6.542)	(8.189)	(7.955)	(5.650)	(5.101)
Observations	12,006	12,006	14,759	14,759	26,765	26,765
R-squared	0.739	0.742	0.744	0.745	0.766	0.768
Fixed Effects	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2
Cluster	GVKEY	GVKEY	GVKEY	GVKEY	GVKEY	GVKEY

Note: We present the results of regression Equation (2). In all columns the dependent variable is the natural log of audit fees. In Columns (1) – (2) we replicate the findings of Fung et al. (2012) for years 2002-2007. In Columns (3) – (4) we extend the analysis to 2008-2018, and in Columns (5) – (6) we analyze both the original sample period and the extended years. All variables are defined in Appendix A. *, **, and *** signify statistical significance at the 10%, 5%, and 1% significance level, respectively.

TABLE 5

Examining Nonlinearity in Audit Fees, Scale, and Specialization

DV=*LAF* VARIABLES	(1) 2002-2007	(2) 2002-2007	(3) 2002-2007	(4) 2002-2018	(5) 2002-2018	(6) 2002-2018
Spec	0.420***	0.373***	0.387***	0.346***	0.305***	0.320***
	(13.71)	(14.70)	(13.59)	(13.69)	(14.78)	(13.71)
Scale	-0.716***	-0.0512	-0.510**	-0.966***	-0.0876***	-0.737***
	(-2.802)	(-1.193)	(-2.114)	(-5.156)	(-2.673)	(-4.246)
$Scale^2$	2.036***		1.546**	2.486***		1.860***
	(2.638)		(2.135)	(4.363)		(3.543)
$Scale^3$	-1.495***		-1.187**	-1.736***		-1.274***
	(-2.614)		(-2.215)	(-4.102)		(-3.278)
*Spec*Scale*	-0.437	-0.401***	-0.299	-0.0900	-0.311***	-0.0841
	(-1.183)	(-8.675)	(-0.868)	(-0.324)	(-8.731)	(-0.326)
$Spec*Scale^2$	-0.394		-0.709	-1.028		-1.022
	(-0.375)		(-0.723)	(-1.310)		(-1.402)
$Spec*Scale^3$	0.479		0.684	0.855		0.862*
	(0.636)		(0.974)	(1.527)		(1.658)
Observations	12,006	12,006	12,006	26,765	26,765	26,765
R-squared	0.743	0.764	0.765	0.769	0.787	0.787
Fixed Effects	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2
Cluster	GVKEY	GVKEY	GVKEY	GVKEY	GVKEY	GVKEY

Note: We present the results of regression Equation (2) with higher order terms. In all columns the dependent variable is the natural log of audit fees. In Columns (1) – (3) we present years 2002-2007. In Columns (4) – (6) we extend the analysis to 2002-2018. All variables are defined in Appendix A. *, **, and *** signify statistical significance at the 10%, 5%, and 1% significance level, respectively.

TABLE 6

Audit Fees, Scale, and Specialization: Parsimonious Models

DV=*LAF* VARIABLES	(1) Table 4 Column (1)	(2) Only *Spec*	(3) Only *Scale*	(4) Only *Citysize*	(5) No *Citysize*
Spec	0.214***	0.181***			0.182***
	(11.87)	(10.20)			(9.972)
Scale	-0.247***		0.0678**		-0.00621
	(-6.285)		(1.967)		(-0.176)
Citysize	0.0845***			0.0596***	
	(12.25)			(9.662)	
LTA	0.495***	0.508***	0.520***	0.510***	0.508***
	(55.37)	(55.52)	(57.63)	(57.68)	(55.67)
LSEG	0.0728***	0.0728***	0.0753***	0.0778***	0.0727***
	(6.003)	(5.905)	(6.025)	(6.330)	(5.868)
CATA	0.321***	0.342***	0.350***	0.320***	0.343***
	(5.070)	(5.327)	(5.445)	(4.986)	(5.365)
Quick	-0.00853*	-0.00877*	-0.00854*	-0.00844*	-0.00877*
	(-1.729)	(-1.794)	(-1.741)	(-1.690)	(-1.794)
DE	0.0325	0.0239	0.0280	0.0406	0.0235
	(0.972)	(0.734)	(0.861)	(1.228)	(0.719)
ROI	-0.0915***	-0.0942***	-0.0973***	-0.0941***	-0.0943***
	(-5.106)	(-5.179)	(-5.138)	(-5.156)	(-5.175)
Foreign	0.236**	0.236**	0.231**	0.225**	0.236**
	(2.158)	(2.114)	(2.121)	(2.140)	(2.109)
Opinion	0.422***	0.418***	0.428***	0.433***	0.418***
	(9.299)	(9.104)	(9.306)	(9.544)	(9.099)
YE	-0.164***	-0.176***	-0.185***	-0.177***	-0.176***
	(-7.490)	(-7.850)	(-8.190)	(-7.990)	(-7.848)
Loss	0.0986***	0.0986***	0.0996***	0.0943***	0.0989***
	(4.930)	(4.864)	(4.869)	(4.667)	(4.875)
AAclients	-0.155	-0.181	-0.178	-0.156	-0.181
	(-1.099)	(-1.263)	(-1.173)	(-1.032)	(-1.265)
Constant	1.007***	2.505***	2.484***	1.402***	2.506***
	(7.251)	(36.99)	(35.99)	(10.49)	(36.41)
Observations	12,006	12,006	12,006	12,006	12,006
R-squared	0.739	0.733	0.728	0.733	0.733
Fixed Effects	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2	Year, SIC2
Cluster	GVKEY	GVKEY	GVKEY	GVKEY	GVKEY

Note: We present the results of alternative specifications of regression Equation (2). In all columns the dependent variable is the natural log of audit fees. All variables are defined in Appendix A. *, **, and *** signify statistical significance at the 10%, 5%, and 1% significance level, respectively.