ABSTRACT

Title of Dissertation: INDUSTRY LINKAGES AND AUDIT FIRMS' INDUSTRY PORTFOLIO CHOICE: EVIDENCE FROM PRODUCT LANGUAGE

Wenfeng Wang, Doctor of Philosophy, 2016

Dissertation directed by: Associate Professor & KPMG Faculty Fellow, Rebecca Hann, Accounting and Information Insurance

Audit firms are organized along industry lines and industry specialization is a prominent feature of the audit market. Yet, we know little about how audit firms make their industry portfolio decisions, i.e., how audit firms decide which *set* of industries to specialize in. In this study, I examine how the *linkages* between industries in the product space affect audit firms' industry portfolio choice. Using text-based product space measures to capture these industry linkages, I find that both Big 4 and small audit firms tend to specialize in *industry-pairs* that 1) are close to each other in the product space (i.e., have more similar product language) and 2) have a greater number of "between-industries" in the product space (i.e., have a greater number of industries with product language that is similar to *both* industries in the pair). Consistent with the basic tradeoff between specialization and coordination, these results suggest that specializing in industries that have more similar product language and more linkages to other industries in the product space allow audit firms greater flexibility to transfer industry-specific expertise across industries as well as greater mobility in the product space, hence enhancing its competitive advantage. Additional analysis using the collapse of Arthur Andersen as an exogenous supply shock in the audit market finds consistent results. Taken together, the findings suggest that industry linkages in the product space play an important role in shaping the audit market structure.

INDUSTRY LINKAGES AND AUDIT FIRMS' INDUSTRY PORTFOLIO CHOICE: EVIDENCE FROM PRODUCT LANGUAGE

By

Wenfeng Wang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Rebecca Hann, Chair
Professor Michael Kimbrough
Professor Oliver Kim
Professor Richmond Mathews
Professor Andrew Sweeting

Acknowledgements

I would like to express my deepest gratitude for my advisor, Dr. Rebecca Hann for her consistent support and mentorship over the past five years. She has helped me to learn not only research but also life. I could not have imagined having a better advisor for my PhD study. I would like to thank my other committee members, Dr. Michael Kimbrough, Dr. Oliver Kim, Dr. Richmond Mathews, and Dr. Andrew Sweeting, for their valuable and constructive suggestions to improve the quality of my work. I would also like to extend my gratitude to all the members of the Accounting and Information Assurance Department at the Robert H. Smith School of Business, notably Dr. Stephen Brown, Dr. Shijun Cheng, Dr. Hanna Lee, Dr. Martin Loeb, Dr. Lawrence Gordon, Dr. Nick Seybert, Dr. Musa Subasi, Dr. Lei Zhou, Dr. Emanuel Zur, Dr. Congcong Li, Dr. Charles Ham, Dr. Kyungran Lee, Yue Zheng, Sijing Wei. Finally I have to say that I could not have my education with the love and support from all my friends and family over the years, especially my wife Le Wang, my father Chengzhong Wang, my mother Haiqin Hao, and my brother Wenjing Wang.

Table of Contents

List of Tables

1. Introduction

A large stream of research in the auditing literature examines how audit market structure affects auditor-client contracting features (e.g., Simunic, 1980, Craswell et al., 1995, Ferguson et al., 2003, Gramling and Stone, 2001, Mayhew and Wilkins, 2003, Francis et al., 2005, Fung et al., 2012, Numan and Willekens, 2013). However, the literature often takes the existing audit market structure as given and we know little about what factors give rise to the existing audit market structure (Defond and Zhang, 2014, Numan and Willekens, 2013).[1] The purpose of this study is to fill this void by examining how audit firms' industry portfolio decisions—their decisions to *jointly* specialize in a set of industries—affect the organization of the audit market. Specifically, I view the audit market as a network of professional services and examine one dimension of this network—the industry linkages in the product space—separately for the "Big 4 audit firms" and the "small audit firms".

Industry specialization is a prominent feature of the audit market and audit firms organize their audit practice along industry lines and hence form their industry-specific expertise (Emerson, 1993).[2] Intuitively, industry connections should play an important role in audit firms' industry portfolio decisions. From the view of an audit firm, while having auditors with industry-specific expertise is necessary to promote higher audit quality and contribute to economic efficiencies, building industry-specific expertise requires substantial investments, which may prevent audit firms from capitalizing on existing resources across multiple industries (e.g., inability to enjoy economies of scale on employee training, transfer human resources, or share

---

[1] Economic theory going back to Hotelling (1929) and Lancaster (1966, 1979) has proposed that a firm's product provision is a strategic and endogenous decision.

[2] Individual auditors are correspondingly industry specialized. To avoid confusion, I will use audit firms to refer to audit firms and auditors to refer to individual auditors.

fixed assets investments across different industries[3]). This gives rise to the basic trade-off between specialization and coordination put forth by Becker and Murphy (1992). To economize on these coordination costs, audit firms may specialize in industries where the industry-specific audit expertise can be more easily transferred from one industry to another. Cremer et al. (2007) study the language used within and across industries and argue that commonality between the languages used across industries reflects lower coordination costs and higher synergistic benefits of organizing across industries for multi-industry firms. They argue that a broader firm, which must communicate and coordinate across industries, is more likely when the degree of language commonality and potential synergies are high. In this study, I apply this theoretical framework in the context of audit market and examine how the linkages between industries in the product space, which are constructed based on textual information of firms' products and reflects the trade-off between specialization and coordination (Cremer et al., 2007), affect audit firms' industry portfolio choices.

In my empirical analysis, I examine three basic spatial characteristics of industry-pairs in the product space: 1) across-industry similarities between two industries that enable lower coordination cost and greater synergistic benefits, 2) the number of industries located between two industries (with more "between-industries" providing greater mobility and entry synergies for audit firms to expand their industry portfolio in the product space), and 3) the degree of within-industry similarity (with higher degree of within-industry similarity more likely causing higher coordination cost across industries but also greater within-industry economies of scale). To construct the product space and across-industry relatedness measures, I follow the approach

[3] It is common that an audit firm provides audit service for multiple industries.

of Hoberg and Philips (2010, 2015), which use computational linguistics to analyze the words firms use in their products descriptions (from firms' 10-Ks filings with the SEC).

My empirical analysis is based on a sample of U.S. audit firms for the period of 2003 to 2013. Because the market structures of the Big 4 and the small audit firms are fundamentally different—Big 4 audit firms are known for their large scale and diverse client bases present in almost all industries and small audit firms tend to have limited capacity and clients in fewer industries, the notion of industry specialization differs across the two markets.[4] For the Big 4 audit market, industry specialization indicates an audit firm's dominant position in the corresponding industry. In contrast, for the small audit firm market, industry specialization reflects more on their industry presence (i.e., which industries they jointly operate in) rather than their dominant position in the industry[5].

For the analysis of Big 4 audit market, I measure Big 4's industry specialization at the three-digit SIC level and identify each firm's industry specialty portfolio (i.e., the set of audit firms' specialized industries). Following Minutti-Meza (2013), I define a Big 4 audit firm as a specialized audit firm in an industry if it is the market leader of the industry and enjoys at least a 10% lead in market share than the second-ranked audit firm. Using a Probit model to test the joint specialization decisions of Big 4 audit firms on industry-pairs, I find that Big 4 firms are more likely to jointly specialize in 1) industry-pairs that are close to each other in the product space, which have lower coordination costs and enjoy greater synergies, 2) industry-pairs with more between-industries, which provide greater mobility and entry synergies for the audit firms to expand their industry portfolio in the product space, and 3) industry-pairs with higher within-

[4] This is often referred to as the "segmented audit market" or "two-tiered audit market" in the literature (Ferguson et al., 2014).

[5] Small audit firms are limited in resources. For any industry, small audit firms will never be a market leader in terms of market share.

industry similarity, which tend to experience greater benefits of within-industry economies of scale.

For the analysis of small audit firm market, an audit firm's industry specialty portfolio is defined as the set of industries it operates in. I find that (similar to Big 4 audit firms) synergistic values between industries also explain small audit firms' industry portfolio choices; however, unlike Big 4, small audit firms are less likely to specialize in highly homogeneous industries (i.e., industries with high within-industry similarity). This is consistent with the interpretation that resource-constrained small audit firms do not have the capacity to take advantage of economies-of-scale benefits and thus stand at a disadvantage in terms of cost efficiencies.

Given the findings that industry specialization decisions of Big 4 audit firms depend not only on an industry's standalone characteristics but also on across-industry linkages in the product space, I further examines whether such across-industry synergies could manifest in auditor-client contracting process in the Big 4 audit market. For this purpose, I use Arthur Andersen's collapse in 2002 as an exogenous shock to the supply side of the market to test how the relative distance between the industries of former Arthur Andersen clients and the Big 4 audit firms' industry specialty portfolios affect the matching between former Arthur Andersen clients and the Big 4. I find that a typical former Arthur Andersen client is more likely to be matched with a Big 4 audit firm that has an industry specialty portfolio more closely aligned with its industry; further, I find that firms with better matches generally enjoy more audit fee discounts. This is consistent with audit firms' strategic specialization decisions in that they offer lower audit fee to attract new clients to either strengthen or expand their existing industry specialty portfolios.

This study makes the following contributions to the literature. First, to the best of my knowledge, this study is the first to empirically investigate how across-industry linkages affect the industry portfolio decisions of audit firms. Gerakos and Syverson (2015) study the implication of potential audit market structure change (i.e., mandatory auditor rotation requirements and further concentration of the Big audit firm market) by empirically modeling auditor-choice decisions of client firms. Their modeling approach emphasizes more on the auditor-choice decisions of client firms (i.e., the demand side of the audit market). By examining how audit firms make their industry-portfolio decisions, this study sheds light on the supply side of the audit market. A better understanding of audit market structure as a whole could help researchers and regulators better assess its potential effect on the efficiency of audit market. Second, this paper adds to the literature on the industry specializations of the Big 4 audit firms. Prior literature has focused more on how within-industry characteristics (e.g., complexity levels of industry-associated auditing processes, degree of regulation in the industry, and homogeneity of firms within an industry) affect audit firms' industry-specialization decisions (Hogan and Jeter, 1999, Cairney and Yong, 2006, Bills et al., 2014). This study complements this literature by exploring how across-industry linkages affect the composition of audit firms' industry specialty portfolios. Third, by examining the industry portfolio choice of audit firms, this study also adds a new dimension to the client-portfolio management literature, which mainly focuses on the effect of audit risk on audit firms' client portfolio management decision (Johnstone and Bedard, 2003, 2004).

The remainder of this paper proceeds as follows. Section 2 discusses related literature and develops research hypotheses. Section 3 discusses data, and construction of main variables. Section 4 presents the empirical results. Section 5 presents results of additional analyses. Section 6 concludes.

2. Hypothesis Development and Related Literatures

2.1 Industry Fundamentals and Firm Organization

The theoretical foundation of my research lies in the literature of the boundary of firms. The trade-off between specialization and coordination affects the scope of firms. Becker and Murphy (1992) theorize the basic trade-off between the cost of coordination and the benefit of efficiency from having workers with different specializations. Intuitively, such trade-off would affect multi-industry firms' industry entry decision. Modeling firms as an information communication network, Bolton and Dewatripont (1994) focus on how certain organization forms could minimize communication costs among different specialized agents within a firm. Relatedly, Hart and Moore (2005) investigate the optimal hierarchy structure inside the firms when there are both specialists and coordinators. Closely related to this study, Cremer et al. (2007) focus on languages that firms use within and across industries and argue that the language commonality across industries affects/reflects the coordination costs and synergistic values between industries. They predict that broader and less specialized product languages can lower the cost and increase the benefits of organizing across industries. Their theory shows how a broader firm, which must communicate across industries, is more likely when the degree of language overlap and potential synergies are high and the cost of imprecise communication (i.e., coordination) is low.

This theoretical formwork has been empirically applied to study the organizational structure of firms. Garicano and Hubbard (2009) study the organization of law firms. They investigate how lawyers with different field specializations (e.g., corporate law, criminal law, and tax law) choose to work together and provide evidence that law firms' field boundaries reflect differences in agency costs related to different fields' cognitive closeness. Specifically, they show that law specialists are more likely to work with others in their own field than those in other fields because of the relatively low costs of monitoring and the benefit of knowledge sharing. In the context of corporate merger and acquisition, Hoberg and Philips (2010) show that firms merge and buy assets to exploit synergies to create new products. It is also found that potential synergies between different industries motivate firms to jointly operate in multiple industries (Hoberg and Philips, 2015).[6]

Applying this framework to the audit market and test the prediction of Cremer et al. (2007), I argue that to the extent that the language used in an industry for its products description reflects the unique production processes, organizational structure, and technology applications, which require different industry-specific auditing expertise, audit firms can capture synergies for industries that share higher degree of language commonality. Such synergies could come from the spillover effect of knowledge across industries (e.g., high mobility to reorganize human resource to other industries), cost sharing from common fixed asset investments (e.g., centralized technical expertise aid) and employment training. Focusing on the consequences of audit firm mergers, Gong et al. (2014) provide evidence that there is significant efficiency increase after consolidation.[7]

2.2 Spatial Characteristics of Industries in the Product Space

In this section, I discuss the key measures that I use to capture across-industry relatedness in the product space, including the utilization of firms' textual information about their products to construct such measures.[8] Product space is defined as a spatial network representation of the relatedness between products in the market. Similar products are closer to each other in the

---

[6] A firm's diversification decision may be motivated by other reasons such as optimal-contracting between shareholders and managers or may reflect the emperor-building behavior of managers. I do not rule out such alternative explanations.

7 Gong et al. (2014) focus on the consequence of audit firm mergers. Questions such as how audit firms merge (e.g., what kind of audit firms choose to merge together?) are out of the scope of Gong et al. (2014).

[8] See section 3 for detailed variable construction information.

product space. Consequently, an industry could be considered a cluster of firms with similar products in the product space. Treating industry as the unit of study, a similar spatial space could be constructed at the industry level. Following Hoberg and Philips (2015), I use the language firms use to describe their products to construct such an industry space. Given any two industries, their spatial locations will define the following industry characteristics.

Across Industry Language Similarity

Across different industries, the distance between two industries could be measured by the extent of product language overlap. The relative closeness of two industries in the product space will reflect the potential of the synergistic value (e.g., a higher degree of relative closeness yields lower coordination costs and higher synergistic value).

Fraction Industries Between Pair

Another unique feature that could reflect the potential synergies of an industry-pair is the extent that other industries locate between the pair. A greater number of industries located between industry-pairs provide more entry opportunity for the audit firm to enter into these between-industries.[9]

Within Industry Language Similarity

Highly specialized industries tend to use more industry-specific, unique language in their business operation (e.g., oil mining industry). An industry's degree of specialization could also affect audit firms' industry portfolio choice. To the extent that more specialized industries have less language commonality with other industries the coordination costs across industries are likely to be higher when an audit firm's client portfolio comprises industries that are highly

[9] I formally define "between industries" in section 4.

specialized. I use the average product language similarity among different firms in the same industry to capture the degree of specialization of an industry.

It is worth to note that such measurement of industry characteristics is not feasible when only relying on traditional SIC or NAICS industry classifications. Though we can roughly say that SIC3 industry-pairs derived from the same SIC2 industry are more similar to each other than SIC3 industry-pairs derived from different SIC2 industries, judgment is hard to make for other cases. For example it is hard to compare two SIC3 industry-pairs derived from the same SIC2 industry. In addition, the text-based measure of industry characteristics is updated yearly (along with firms' annually filed10-Ks). In contrast, SIC or NAICS industry classifications change rarely; thus, they can better capture the variation of across-industry characteristics (Hoberg and Philips, 2015).

2.3 Segmented Audit Market

The U.S. audit market, associated with publicly registered firms, is a typical example of a segmented market (Francis and Simon, 1987) and is also referred to as a two-tiered market (Ferguson et al., 2014). For large registrants, the audit market is an oligopoly formed by the Big-N audit firms.[10] The market share of Big N auditors, measured by the total assets of client firms, has been consistently over 95% over the last decade. For the fiscal year 2014, Big 4 provided audit services to 3,061 big public firms while more than 400 small audit firms provided audit services to the other 4,032 relatively small public firms (Audit Analytics, 2015). There are systematic differences in the audit capabilities of Big 4 and non-Big 4 audit firms (Ferguson et al., 2014). The relative small scale of the small audit firms limits their ability to audit large

[10] Over time, the U.S. big public firms' audit service has concentrated more and more in relatively fewer large audit firms. Before 2002, there were initially Big Five audit firms (i.e., Arthur Andersen, Deloitte, EY, PwC, and KPMG); however, after 2002 (with the collapse of Arthur Andersen), the so-called Big 4 emerges (i.e., Deloitte, E&Y, PwC and KPMG).

public firms. Thus, I develop my research hypotheses and conduct empirical analyses separately for the two markets.

3. Hypotheses

In this section, I develop my research hypotheses with respect to each of the across-industry relatedness measures I am investigating.

Across Industry Language Similarity

Ex ante, there are both benefits and costs for Big 4 audit firms to specialize in industries close to each other in the product space. On the one hand, Big 4 have incentive to specialize in similar industries due to potential synergies (Carter, 1977, Hoberg and Philips, 2015), the benefits of economies of scale, and lower coordination costs (Becker and Murphy, 1992, Bolton and Dewatripont, 1994, and Cremer et al., 2007). On the other hand, because of their deep-pocket Big 4 face high litigation risk (Amihud and Lev, 1981, Gilson and Mnookin, 1985). As a result it is also likely that they tend to form a more diversified industry specialty portfolio to mitigate the potential litigation cost. This argument does not apply to small audit firms as they typically have less litigation risk (Hope et al, 2010). In addition, also due to their limited size small audit firms are not able to fully economize on fixed costs as Big 4 audit firms do. In this sense the marginal benefit of synergies across industries is likely to matter more for small audit firms than for Big 4 audit firms. Hence, my first two hypotheses are:

H1.1: Ceteris paribus, the industry specialty portfolio choices of Big 4 audit firms are not related to Across Industry Language Similarity.

H1.2: Ceteris paribus, small audit firms are more likely to jointly specialize in industry-pairs with high Across Industry Language Similarity.

Figure 1-A depicts that, in a simplified two-dimensional space, industry X and Y are close to each other and thus have high-across industry synergies and lower coordination costs than other feasible industry-pairs. *H1.2* predicts that small audit firms are more likely to jointly specialize in industry X and Y than other possible industry-pairs, such as the pair comprised of Industry 1 and 2 or the pair comprised of Industry X and 2.

Fraction Industries Between Pair

Fraction Industries Between Pair is defined as the fraction of industries located between a given industry-pair in the product space. Synergies from between-industries come from the potential entry opportunities. *Figure 1-B* depicts this case where Industry 2 is a between-industry of Industry X and Y.[11] For an audit firm, jointly specializing in Industry X and Y provides the firm advantages to enter into Industry 2. Such potential entry opportunities are valuable for both Big 4 audit firms and small audit firms. As a result, I have directional hypothesis for both Big 4 audit firms and small audit firms.

H2.1: Ceteris paribus, Big 4 audit firms are more likely to jointly specialize in industry-pairs with more between-industries.

H2.2: Ceteris paribus, small audit firms are more likely to jointly specialize in industry-pairs with more between-industries.

Within Industry Language Similarity

Within industry language similarity captures the homogeneity of firms within the same industry. Firms in such industries are less spread-out in the product space. *Figure 1-C* depicts Industry X and Y as having low degrees of within industry language similarity compared to other

[11] Please refer to section 3.3 for more refined definition of between-industry.

industries. Such industries tend to be more specialized industries such as the oil industry and mining industry. Within industry language similarity has two differential effects on audit firms' industry portfolio choice. On one hand, more specialized an industry is, the higher the coordination cost is if organized with other industries. On the other hand, high similarity among firms within the same industry also implies high benefits of within-industry economies of scale. The two opposing effects seem to make the net effect on audit firms' industry portfolio choice unclear. I argue that Big 4 audit firms stand at a better position to take advantage of economies of scale due to their capacities to provide services to many clients in the same industry and they are more likely to jointly specialize in industry-pairs with high within industry language similarity. The opposite is true for small audit firms.

H3.1: Ceteris paribus, Big 4 audit firms are more likely to specialize in industry-pairs with high within industry language similarity.

H3.2: Ceteris paribus, small audit firms are less likely to specialize in industry-pairs with high within industry language similarity.

4. Related Literature

Strategic Industry Choice Decision of Audit Firms

As in Gerakos and Syversion (2015), I consider the audit market as a differentiated product market in which audit firms provide differentiated audit services and client firms rationally choose the audit firm that provides them with highest utility. The market equilibrium is determined by both the supply side of the market (i.e., audit firms' product provision decisions: which industries, audit quality, and audit price) and the demand side of the market (i.e., client firms' specific demand for audit quality, willingness to pay, and preferences for different audit firms).[12] Theoretically, in a simplified one period model, this could be modeled as a process in which audit firms first make their product provision decisions, and then given the demand side characteristics the market clears at negotiated audit price. Such modeling approach has been used in existing theoretical research. Chan et al. (1997) study a one period spatial auditing competition model in which each of the audit firms chooses multiple industries to specialize in. Additionally, Chan (1999) examines a setting in which audit firms are ex-ante identical and strategically choose to become ex-post industry specialized. In this study, I specifically focus on and examine how across-industry relatedness affects audit firms' strategic industry portfolio choice.

Why Do Audit Firms Specialize?

For an oligopoly market, product differentiation is firms' optimal product choice. Competition level among firms could be soften through product differentiation (Mazzeo, 2002). For the Big 4 audit market, audit firms could differentiate themselves by specializing in different industries and obtain competitive advantage in their specialized industries (Johnson and Lys, 1990).

Industry specialization is a prominent feature of modern audit practice (Bell et al., 1997). Audit firms are organized along industry lines and designate most, if not all, of their auditors as industry specialists (Emerson, 1993). Existing empirical studies on the industry specialization decisions of audit firms mainly focus on the question of why we observe an specialized audit firm in certain industries but not in other industries. Beginning with studies by Eichenseher and Danos (1981) and Danos and Eichenseher (1982, 1986), prior studies showed that homogeneity of firms within the same industry and the associated auditing complexity are significant determinants of audit firms' industry specialization decisions. For the sample period from 1976

[12] Audit firms' preference over potential clients' characteristics could be reflected in the audit price and the decision of the audit firm to bid for certain potential client's audit engagement.

to 1993, Hogan and Jeter (1999) find that while industry specialization remained greater in regulated industries, specialization has also been observed for non-regulated industries over time. Cairney and Young (2006) and Cahan et al. (2008) document that audit firms are more likely to specialize in industries with greater homogeneity among clients' operations and investment opportunity sets. The central argument in these studies is that economies of scale exist in industries of greater homogeneity in which audit expertise could be more easily transferred from one firm to another. My study is different from this literature and adds to it in that I examine how audit firms choose to specialize in a set of industries rather than one particular industry and I approach the question from the perspective of across-industry linkages.

Audit Firms' Client Portfolio Management

My study is also related to the literature on audit firms' client portfolio management. In considering how audit firms decide whether or not to accept potential new clients, Simunic and Stein (1990) propose that, to a given firm, "the riskiness of an audit engagement is properly defined and measured within a portfolio context rather than in isolation." In other words, they argue that new clients' engagement risk depends not only on the characteristics of the prospective pool of potential clients but also on the characteristics of an audit firm's existing client portfolio. Shu (2000) show evidence that audit firms actively manage their client portfolio and would resign from misaligned clients. Specifically, the alignment studied in Shu (2000) refers to the matching between big vs. small audit firms and big vs. small client firms. Johnstone and Bedard (2004) empirically compare the relative riskiness of newly accepted firms, retained firms, and dropped firms. Taking the portfolio view as in Simunic and Stein (1990) and applying it to audit firms' industry portfolio choice, I argue that audit firms' industry portfolio decisions depend not only on the industries' standalone characteristics but also on the across-industry relatedness. A closely related and interesting study is Brown and Knechel (2013) where they examine how auditor-client compatibility affects the matching results between audit firms and clients. The "compatibility" measure they use is defined as "similarity between a client firm and an audit firm's existing clients in the same industry". Similar to early studies (i.e., Hogan and Jeter, 1996, and Cairney and Young, 2006), Brown and Knechel (2013) focuses on audit firms' within industry client portfolio management. My study differs from Brown and Knechel (2013) in that I examine audit firms' industry portfolio choice which emphasizes on

5. Data and Construction of Main Variables

5.1 Main Datasets Used

To conduct my empirical analysis, I used three main data sources. Firm-level financial variables are from COMPUSTAT Industrial Annual Files provided by WRDS. Data on audit firms are from Audit Analytics. To construct the product space, I use the "10-K Text-based Network Industry Classifications Data" (TNIC) provided by the Hoberg-Philips Industry Classification Library.[13] I use the "Use Table" of Benchmark Input-Output Accounts of the U.S. Economy to measure vertical relatedness between industries. My sample period is from 2001 to 2013. I provide main data selection procedure here. Some other data selection processes will be addressed when I discuss variable construction or empirical model specification.

1. Since I examine the U.S. audit market, foreign incorporated firms are dropped (Gul et al., 2009, Minutti-Meza, 2013).
2. Firm-years without specific SIC industry classification are dropped. Such cases include missing SIC variable or unidentified SIC (i.e., SIC = 9999 or 0000).
3. For cases in which a firm has multiple auditors, I only keep the main auditor whose audit fee is over 70% of the total audit fee.[14]

[13] Gerard Hoberg and Gordon Philips share their data through their website at http://cwis.usc.edu/projects/industrydata/industryclass.htm.

[14] Compustat database only records firms' main auditor. Audit Analytics database records all the auditors providing audit service to the firm in the same year.

5.2. Construction of the Product Space

In this section I describe the construction process of the product space based on the textual information of firms' product description in their 10-Ks. The basic building block of the construction of product space is a binary words vector representing each firm's product description. To build such word vectors, Hoberg and Phillips (2010) use web crawling and text parsing algorithms to construct a database of product descriptions from firms' 10-K annual filings. Product descriptions are legally required to be accurate, and these descriptions must be updated and representative of the current fiscal year of the 10-K. Using firms' product descriptions, they form word vectors for each firm. After excluding trivial words often used in product descriptions (e.g., the, a, we, use etc.), each firm is assigned a column vector representing the firm's location in the product space. For example, suppose there are M_t non-trivial words at fiscal year t. For a firm j in year t, its word vector $W_{j,t}$ is a binary M_t vector, having value one for a given element when firm j uses the given word in its fiscal year t 10-K product description. Then each firm's word vector is normalized to unit length, resulting in the normalized word vector $N_{j,t}$. In this way each firm is represented by a unique vector of length one in a M_t dimensional product space.[15] Therefore, all firms reside on a M_t unit sphere, and each firm has a known location. This spatial representation of the product space allows for constructing variables that more richly measure industry topography to identify, for example, the distance between two industries, and other industries that lie between a given pair of industries.

[15] In *Appendix B*, for simplicity I provide examples in a two dimensional space.

Cosine similarity is used to measure the distance between any two points in the M_t dimensional space.[16] The cosine similarity for any two word vectors $N_{j,t}$ and $N_{i,t}$ is the vector dot product $N_{j,t} \cdot N_{i,t}$. The cosine similarities are naturally bounded between zero and one. If two firms have exactly the same product description, their dot product will tend towards one while dissimilarity moves the cosine similarity toward zero.

$$Similarity_{\mathrm{ijt}} = N_{j,t} \cdot N_{i,t} \qquad (1)$$

5.3 Main Variables Construction

5.3.1 Text-based Industry Variables

Across Industry Language Similarity (AILS)

This measure is based on industry product language overlap and captures the extent to which product descriptions of firms in two different industries use overlapping language. High-similarity level indicates high-synergistic benefit of jointly operating in the two industries. It is constructed as the average textual cosine similarity of all pairwise permutations of the N_i and N_j firms in the two SIC3 industries i and j in a given year, in which textual similarity is based on word vectors from firm product description as discussed above.

Between-Industry

Given the across industry language similarity measure for any two SIC3 industries in the product space, it can be used to assess which other industries lie between two SIC3 industries. A third industry is defined to be between two industries in a given industry pair if the third industry is closer in textual distance to each industry in the pair than the two industries in the pair are to each other.

---

[16] Cosine similarity method is widely used in studies of information processing. See Sebastiani (2002) for a summary of the method.

Let $AILS_{i,j}$ denotes the across-industry product language similarity of industries i and j, a third industry k is defined as being between industries i and j and is called a between-industry if the following conditions are met:

$$AILS_{k,i} \geq AILS_{i,j} \quad and \quad AILS_{k,j} \geq AILS_{i,j}$$

Fraction Industries Between Pair

The fraction of industries between a given pair of industries i and j is defined as the number of industries k satisfying this condition divided by the total number of industries in a given year.

Within Industry Language Similarity (WILS)

Within industry language similarity is measured as the average cosine similarity of the product descriptions for all pairwise permutations of the N_i firms in industry i in a given fiscal year. It captures the homogeneity level of firms within the same industry. More unique and specialized industries would tend to have high within industry language similarity.

5.3.2 Non-text-based Industry Control Variables

Since I examine how across-industry relatedness in the product space affects industry choice of audit firms, it is important to control for other possible confounding factors. In my analysis, I control for the following potential confounding factors.

Vertical Relatedness

Many prior studies have examined the effect of industries' vertical relatedness on corporate policy and structure (Fan and Goyal, 2006, Ahern and Harford, 2014). It is likely to affect audit firms' industry portfolio choice. For example, audit demand of vertically integrated firms could motivate audit firms to specialize in the corresponding vertically related industries. I use Bureau of Economic Analysis's industry Input-Output tables to measure the degree to which two SIC3 industries are vertically related. I follow the methodology described by Fan and Goyal (2006) to define the vertical relatedness of two SIC3 industries.

Pair Likelihood if Random

If audit firms' industry portfolios are randomly formed, industries with more firms are more likely to be observed in an audit firm's industry portfolio. Since I will test the likelihood of an industry pair belonging to the same audit firm's industry portfolio, it is necessary to control for industry pair relevance. For example, if audit firms randomly choose client firms, then the probability of two industries belonging to the same audit firm's industry portfolio would be positively associated with the product of the fraction of firms residing in the two industries. I define *Pair Likelihood if Random* as the product $F_i \cdot F_j$, where F_i is the number of firms in industry i divided by the total number of firms in the market in the given year (F_j is similarly defined).

Economies of Scale

Following Hoberg and Philips (2015), I also consider economies of scale and measure the gains to scale within each industry. This is measured by estimating a translog Cobb-Douglas production function. Ten years of lagged data for each firm in a given SIC3 industry is used. I use *sales* as the dependent variable of the estimation. The independent (input) variables include: *net property plant and equipment* for capital, *the number of employees*, *the cost of goods sold*, and also the *age of the firm*. All variables are in natural logs, and variables except for *age* and the *number of employees* are deflated to 1987 real dollars using the wholesale price index. With the estimated production function, an industry's economy of scale is defined as the sum of the coefficients on net property plant and equipment and the cost of goods sold. For a pair of SIC3 industries, *Economies of Scale* is defined as the mean of the two industries' economies of scale.

Industry Stability

It is likely that audit firms would avoid specializing in more volatile industries. For this concern, I also control for Industry Instability in my analysis. I compute each SIC3 industry's instability as the absolute value of the natural logarithm of the number of firms in the industry in year t divided by the number of firms in the same industry in year $t-1$. Industries with higher instability are experiencing changes in the industry's membership over time. For a given SIC3 industry pair, *Industry Stability* is defined as the mean of the two industry's stability.

Industry Size

Intuitively, industry size (i.e., the number of firms in the industry) is a factor which would affect audit firms' ability to take the benefit of economies of scale. I control the average industry size of two SIC3 industries in an industry pair. I use the number of firms in a given SIC3 industry-year to measure industry size. The variable *Industry Size* is defined as the decile rank of the average size of two SIC3 industries.

Same SIC2

Since SIC3 industry-pairs within the same SIC2 industry group are apparently more similar than SIC industry-pairs composed by two SIC3 industries from different SIC2 industries, to mitigate the concern that the text-based measure I use merely captures whether two SIC3 industries belong to the same SIC2 industry, I also control such variation. Dummy variable *Same SIC2* takes the value of one if two SIC3 industries in a pair belong to the same SIC2 industry, and otherwise takes the value of zero.

5.3.3 Audit Firm's Industry Portfolio

Big 4's Industry Specialty Portfolio

Following prior studies, I identify the Big 4's industry specialization by their within-industry market share. For each audit firm and year, a SIC3 industry market share is calculated as follows:

$$MarketShare_{kit} = \frac{\sum_{j=1}^{J} S_{kijt}}{\sum_{i=1}^{I}\sum_{j=1}^{J} S_{kijt}} \qquad (2)$$

Where $MarketShare_{kit}$ is the market share of audit firm i in SIC3 industry k in year t, S_{kijt} represents the total assets/sales of client firm j in industry k audited by audit firm i in year t, J represents the number of clients that are served by audit firm i in industry k in year t, and I is the total number of audit firms in industry k in year t. An audit firm is defined as a specialist in industry k at year t if it has the largest market share in that industry year and has more than 10% market share lead than the second ranked audit firm.[17] To properly measure audit firms' market share, I drop SIC3 industry-years with less than eight firms. All the industries in which one audit firm specializes in form its industry specialty portfolio in that year.

For a given year t, all the SIC3 industries that have a specialized audit firm form the complete set of industries in which one audit firm could specialize in.[18] Then all the feasible pairwise permutations of the SIC3 industries in the above identified set are the potential industry-pairs one audit firm could specialize in year t. I define a binary variable $Industry\ Pair$ which takes value

[17] I also require that the audit firm has at least 20% of market share in the industry. Dropping this constraint does not change the result.

[18] For clarification, I call these SIC3 industries as specialized industries.

of one if the industry pair belongs to the same auditor's industry specialty portfolio in year t, otherwise takes of the value of zero.

Small Audit Firm's Industry Specialty Portfolio

In the analysis of small audit firms' industry portfolio choice, for each year t I identify their industry portfolios through their client portfolio. I exclude client firms with audit fees less than $5,000 (i.e., more likely to be trivial none-major audit service). I also drop client firms in industries that are not covered in *Hoberg-Philips Industry Classification Library*, because for these industries their product space location is not available.

Other variables used in the analysis are consistent with the literature and are not elaborated here. The variable definitions are provided in Appendix A.

6. Summary Statistics and Empirical Analysis

6.1 Summary Statistics

Table 1 reports summary statistics for audit firms for the period 2001 to 2013. Since I use both Compustat and Audit Analytics data, my analysis covers industries appeared in both datasets. This step drops many investment funds firms reported in Audit Analytics, but are not included in Compustat. I also exclude firm-year observations of the following types of firms: non-U.S. incorporated ones, those with zero or negative audit fees, those with unidentified SIC3 industries (i.e., with SIC3 industry code "000" or "999"), or those in three-digit SIC industries ending with "zero," which are actually two-digit SIC industries. *Table 1* shows a clear pattern of segmented audit market where Big 5 audit firms audited an average of 1,078 large client firms per year.[19] A lot more small auditors provide audit service to relatively smaller public firms. Even the "Second Tier" audit firms are much smaller than the Big 5.[20] The average number of clients of a second-tier audit firm is about one fifth of the average number of clients of the Big 5 audit firms. Taking into account the size difference of the client firms, the scale difference between Big 4 and non-Big 4 audit firms would be even more dramatic. Not surprisingly, the industry operation of small audit firms concentrates on much less number of SIC3 industries compared to Big 5 audit firms. In my sample period, a Big 5 audit firm covers, on average, about 167 different SIC3 industries in a typical fiscal year.

6.2 Big 4's Industry Specialty Portfolio Choice

Table 2 Panel A shows the distribution of Big 4 audit firms' market share in their specialized SIC3 industries for the ten-year period from 2003 to 2013. Since Arthur Andersen collapsed in 2002, I start my sample period from 2003. The market share is calculated basing on the total assets of audit firms' client firms. The mean market share is 60%, implying that the Big 4 do specialize in some industries and have a dominant role their specialized industries. *Table 2 Panel B* reports the distribution of the number of SIC3 industries in which there is one specialized audit firm over time and across the Big 5 audit firms. For example, PwC specialized in 33 different SIC3 industries for fiscal year 2004. Among the Big 4, KPMG is relatively small in terms of the number of specialized SIC3 industries. The total number of specialized SIC3 industries each year ranges from 89 to 115.

[19] For the year 2001, it was still the "Big Five" accounting firms. Arthur Andersen collapsed in 2002.

[20] Recently, Hogan and Martin (2009) consider BDO Seidman, Crowe, Chizek and Company, Grant Thornton LLP; and McGladrey & Pullen as second-tier audit firms. There is not yet well-accepted definition of second tier audit firms. In my sample, I define an audit firm as a second-tier audit firm if its number of clients is between 100 and 500.

Table 1: Descriptive Statistics of Audit Firms

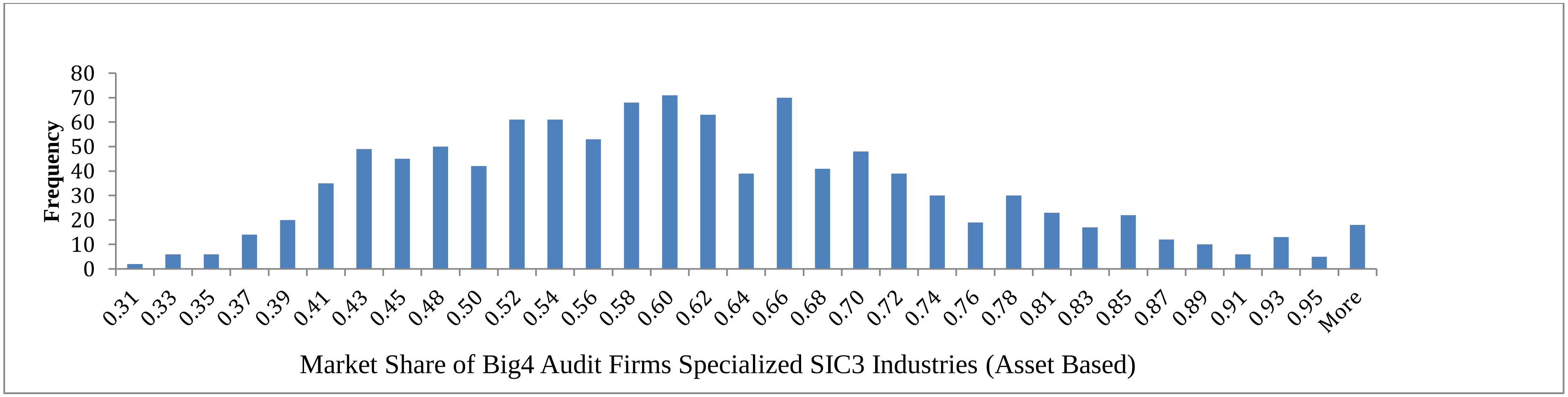
Audit Firm Size	Number of Clients	Obs.	%	Number of Audit Firms	Number of Client Firms		Number of SIC Industries		Audit Fees (Thousands)		Audit Fee Industry Concentration		
					Mean	Median	Mean	Median	Mean	Median	HHI	CR2	CR3
Very Small	1	2,446	0.38	981	1	1	1	1	72	44	1.00	1.00	1.00
Relatively Small	2 - 20	3,556	0.55	770	5	4	4	3	450	236	0.46	0.80	0.89
	21 - 50	311	0.05	86	30	27	15	17	3,687	2,373	0.27	0.51	0.59
	51 - 100	66	0.01	18	69	65	22	24	9,407	8,461	0.29	0.56	0.64
Second Tier	101 - 500	38	0.01	7	212	212	68	80	97,276	90,846	0.09	0.28	0.34
Big N	> 500	53	0.01	5	1,078	1,099	167	166	1,793,331	1,948,178	0.03	0.14	0.19

This table shows the summary statistics for audit firms reported in Audit Analytics for the period 2001 to 2013. My sample starts with the merged results between Compustat and Audit Analytics.[21] I exclude firm-year observations that are non-US incorporated firms, have zero or negative audit fees, are in SIC3 industries "000" or "999" (unidentified industries), or are in SIC3 industries ending with "zero" (these are firms that could not be classified at the three-digit industry level).

[21] This step drops many investment funds which are required to be audited but do not file annual reports.

Table 2: Big 4 Audit Firms' Industry Specialization

Panel A: Descriptive Statistics of Big 4 Audit Firms' Industry Specialization



Panel B: Distribution of Big 4 Audit Firms' Specialized SIC3 Industries

Auditor	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
PwC	40	36	34	33	29	30	30	30	29	30	27	31	29	408
EY	25	34	35	34	30	30	31	28	29	27	22	23	25	373
Deloitte	17	31	25	27	30	31	29	28	26	29	28	22	23	346
KPMG	10	14	15	13	13	15	11	11	12	13	12	13	16	168
AA	17	0	0	0	0	0	0	0	0	0	0	0	0	17
Total	109	115	109	107	102	106	101	97	96	99	89	89	93	1,312

Panel A shows the distribution of Big 4 audit firms' market share in their specialized SIC3 industries. I define one audit firm as a specialized audit firm in one SIC3 industry if the audit firm is the market leader in the industry and has at least 10% lead of market share (assets based) to the second ranked audit firm. The market share is based on the client firms' total asset. Panel B shows the distribution of specialized SIC3 industries over the Big-N firms (PwC: PricewaterhouseCoopers, EY: Ernst & Young, Deloitte: Deloitte Touche Tohmastsu Limited, KPMG, and AA: Arthur Andersen) and over the period 2001 – 2013. Similar descriptive statistics by using client firms' total sale to assess audit firms' industrial specialization is provided in Appendix C.

I use the following Probit model to test my hypotheses on how across industry relatedness affects Big 4 audit firms' industry portfolio choice.

$$\begin{aligned} Pr\left(Industry\ Pair_{ij,t} = 1\right) \\ &= F(\beta_0 + \beta_1 Across\ Industry\ Language\ Similarity_{ij,t} \\ &+ \beta_2 Fraction\ Industries\ Between\ Pair_{ij,t} \\ &+ \beta_3\ With\ Indusry\ Language\ Similarity_{ij,t} \\ &+\ \beta_4\ Vertical\ Relatedness_{ij,t} + \beta_5 Industry\ Stability_{ij,t} \\ &+ \beta_6 Economies\ of\ Scale_{ij,t} + \beta_7 Industry\ Size_{ij,t} + \beta_8 Same\ SIC2_{ij,t} \\ &+ \boldsymbol{Year}) \qquad (3) \end{aligned}$$

To control for potential time trend effect, I control for year fixed effect in the model.[22] Please refer to Appendix A for detailed variable definitions.

Table 3 Panel A reports summary statistics for the variables used for the industry portfolio choice test of Big 4. In the final sample, I have 41,791 SIC3 industry-pairs which are all the feasible pairwise permutations of specialized SIC3 industries for the sample period from 2003 to 2013. Among the 41,791 feasible industry-pairs, 26.5% of the pairs are in the same Big 4 audit firm's industry specialty portfolio. The average of *Across Industry Language Similarity* of all randomly formed SIC3 pairs is 0.0071, which is significantly lower than the average of SIC3 industry-pairs formed within the same Big 4 audit firm's industry specialty portfolio (0.0084), which provides the preliminary evidence that Big 4's industry specialty portfolio decisions are affected by the across-industry relatedness in the product space.

[22] The model does not use year-clustered standard error, since there are only 11 clusters in the sample and small number of clusters would cause biased estimates. As a result, I report robust standard error. However, using year-clustered error does not change my results, and it provides even sharper estimates, indicating only a minor small number of cluster bias issue.

Table 3: Big 4 Audit Firms' Industry Specialty Portfolio Choice

Panel A: Industry Pairs Descriptive Statistics

Variables	Count	Mean	Std. Dev.	Median	P75	Min	Max
Industry Pair	41791	0.265	0.441	0.000	1.000	0.000	1.000
Across Industries Language Similarity	41791	0.007	0.041	0.000	0.000	0.000	0.871
Within Industry Language Similarity	41791	0.346	0.192	0.336	0.475	0.000	1.000
Fraction Industries Between Pair	41791	0.052	0.033	0.046	0.068	0.004	0.352
Vertical Relatedness	41791	0.012	0.031	0.003	0.012	0.000	1.000
Industry Stability	41791	0.065	0.044	0.059	0.091	0.000	0.350
Economies of Scale	41791	0.822	0.125	0.847	0.913	0.177	1.146
Same SIC2	41791	0.023	0.150	0.000	0.000	0.000	1.000
Industry Size	41791	0.492	0.324	0.444	0.778	0.000	1.000

This table reports the summary statistics of the variables used for the analysis of Big 4 audit firms' industry specialty portfolio choice for the sample period 2003 to 2013. The construction of these variables are discussed in detail in section 3. *Across Industry Language Similarity* is the average pairwise similarity of firms in one of the industry in the pair with firms in the other industry. *Within Industry Language Similarity* is the average of the within industry product language similarity of the two SIC3 industry in a pair. *Fraction of Industries Between Pair* is the fraction of all other industries that are located between the given pair of industries. *Vertical Relatedness* is the degree of vertical relations based on the input-output tables. *Industry Stability* is the average of the absolute values of the logarithmic change in the number of firms in the given two industries in a pair over the previous year. *Economies of Scale* is based on the estimation of a Cobb-Douglas production function over ten years. *Same SIC2* is a binary dummy variable which takes value of one if two SCI3 industries are in the same SIC2 industry and otherwise takes the value of zero. *Industry Size* is the decile rank of the average number of firms in the two industries.

Table 3 Continued

Panel B: Pearson Correlation

Variables	Industry Pair	Across-industry Language Similarity	Within Industry Language Similarity	Fraction Industries Between Pair	Vertical Relatedness	Industry Stability	Economies of Scale	Same SIC2
Industry Pair	1.000							
Across Industry Language Similarity	0.013**	1.000						
Within Industry Language Similarity	0.017***	0.146***	1.000					
Fraction Industries Between Pair	0.005	-0.133***	-0.099***	1.000				
Vertical Relatedness	-0.016**	0.099***	-0.019***	-0.012*	1.000			
Industry Stability	0.012*	-0.034***	-0.111***	-0.022***	-0.008	1.000		
Economies of Scale	-0.007	0.013**	0.027***	-0.059***	0.097***	0.014**	1.000	
Same SIC2	-0.019***	0.311***	-0.054***	-0.044***	0.053***	-0.014**	-0.014**	1.000
Industry Size	-0.005	-0.019***	0.126***	0.241***	-0.043***	-0.111***	-0.173***	-0.004

Pearson Correlation Coefficients are reported for my sample of 41,791 observations of SIC3 industry-pairs from 2003 to 2013.

Table 3 Panel B displays the Pearson correlation coefficients for the industry-pair variables. The key variable I examine in the next section is *Industry Pair* which is a binary dummy variable indicating whether an SIC3 industry pair is in the same auditor's industry specialty portfolio. The first column of this table shows that *Industry Pair* is positively associated with *Across Industry Language Similarity*, *Within Industry Language Similarity*, and *Fraction Industries Between Pairs.*

Table 3 Panel C reports the estimate of the Probit model (3). In this model, I test whether potential synergies across industries affect Big 4's choice of industry specialty portfolios. Each observation is a pair of SIC3 industries derived from the set of all pairings of specialized SIC3 industries in a given fiscal year. In model 1, the industry specialty of an audit firm is identified based on the market share of client firms' total assets. In model 2, the industry specialty of an audit firm is identified based on the market share of client firms' total sales.

The results show that Big 4 audit firms are more likely to jointly specialize industry pairs with high *Across Industry Language Similarity*. This confirms the result observed in the univariate correlation table. In my hypothesis development section, I argued that higher litigation costs could motivate the Big 4 audit firms to have diversified industry specialty portfolios (i.e., industries far away in the product space). The results support the view that the negative spillover effect from litigation does not differentiate industries close to or far away from the focal industry or the benefit of across industry synergies exceeds the potential incremental litigation cost.

Consistent with hypothesis *H2.1*, the fraction of industries between an industry pair is positively associated with the probability of the same Big 4 audit firm jointly specializing in the two industries. This is consistent with the interpretation that benefit of opportunities to expand their industry specialties to *Between Industries* motivate Big 4 audit firms to jointly specialize in industry pairs with more *Between Industries*.

Related to hypothesis *H3.1*, the result shows that *Within Industry Language Similarity* of two industries is positively associated with the probability that the same audit firm will jointly specialize in the two industries. This is consistent with the view that, for Big 4 audit firms, the benefits of economies of scale exceed the potential costs related to high coordination costs between specialized industries and other industries. This is also consistent with the finding of Bills et al. (2014) who find that audit firms enjoy cost efficiency in specialized industries and such cost efficiency could be passed to client firms in terms of low audit fee.

Table 3 Continued

Panel C: What Industry-Pairs Big 4 Audit Firms Specialize In

		Model 1	Model 2
	Expected Sign	Coefficient	Coefficient
Across Industry Language Similarity	+ / - ?	0.737***	0.578***
		(4.314)	(3.415)
Fraction Industries Between Pair	+	0.894***	0.665**
		(3.368)	(2.527)
Within Industry Language Similarity		0.117***	0.115***
		(3.277)	(3.15)
Vertical Relationship		-0.769***	-0.824***
		(-3.563)	(-4.048)
Industry Stability		0.424***	0.398**
		(2.804)	(2.522)
Economies of Scale		-0.072	-0.191***
		(-1.322)	(-3.530)
Industry size		-0.054**	-0.114***
		(-2.412)	(-4.976)
Same SIC2		-0.227***	-0.214***
		(-4.608)	(-4.243)
Constant		-0.673***	-0.521***
		(-11.79)	(-9.234)
Observations		41,791	41,308
Year Fixed		YES	YES
Robust Std. Error		YES	YES
Pseudo R-squared		0.0017	0.0022

This table reports the coefficients and associated t-statistics (in parentheses) for the following model:

$$
\begin{aligned}
Pr\left(Industry\ Pair_{ij,t} = 1\right) \\
&= F(\beta_0 + \beta_1 Across\ Industry\ Language\ Similarity_{ij,t} \\
&+ \beta_2 Fraction\ Industries\ Between\ Pair_{ij,t} \\
&+ \beta_3\ With\ Indusry\ Language\ Similarity_{ij,t} + \ \beta_4\ Vertical\ Relatedness_{ij,t} \\
&+ \beta_5 Industry\ Stability_{ij,t} + \beta_6 Economies\ of\ Scale_{ij,t} + \beta_7 Industry\ Size_{ij,t} \\
&+ \beta_8 Same\ SIC2_{ij,t} + \boldsymbol{Year}) \qquad (3)
\end{aligned}
$$

One observation is one pair of Big 4 specialized SIC3 industries in a year derived from the set of all permutations of feasible pairings. The dependent variable is a binary dummy variable which takes the value of one if the two SIC3 industries are from the same audit firm's industry specialty portfolio, otherwise takes the value of zero. In model 1, Big 4's industry specialty is identified according to the market share of their clients' total assets. In model 2, Big 4's industry specialty is identified according to the market share of their clients' total sales. The model is estimated with year fixed effects and robust standard errors for the sample period from 2003 to 2013.

6.3 Small Audit Firm's Industry Portfolio Choice

In this section, I test my hypotheses on the link between small auditors' industry portfolio choice and across-industry relatedness in the product space. Recall that different from the case of Big 4 audit firms, I identify a small audit firm's industry specialty portfolio as the set of SIC3 industries it jointly operates in. This is reasonable in the sense that the limited capacity of small audit firms forces them to operate only in industries for which they have the required audit specialty. Different from Big 4 audit firms, small audit firms' industry specialties do not indicate their dominant market position. For the Big 4 audit firms, one SIC3 industry-pair could be in at most one Big 4 audit firm's industry specialty portfolio. For the small audit firms, however, more than one firm could operate in a given SIC3 industry-pair. To capture the overall tendency of small audit firms to jointly operate in a given industry-pair, I use the ratio: number of small audit firms that jointly operate in a given SIC3 industry-pair divided by the total number of small audit firms in a given year, *Fraction Audit Firms in Existing Industry Pair*. I use the following OLS model to test my hypotheses relate to small audit firms.

$$
\begin{aligned}
&Fraction\ Audit\ Firms\ in\ Existing\ Industry\ Pair_{ijt}\\
&\quad = \beta_0 + \beta_1 Average\ Across\ Industry\ Language\ Similarity_{ij,t}\\
&\quad + \beta_2 Fraction\ Industries\ Between\ Pair_{ij,t}\\
&\quad + \beta_3\ With\ Indusry\ Language\ Similarity_{ij,t} +\ \beta_4\ Vertical\ \ Relatedness_{ij,t}\\
&\quad + \beta_5 Industry\ Stability_{ij,t} + \beta_6 Pair\ Likelihood\ if\ \ Random_{ij,t}\\
&\quad + \beta_7 Economies\ of\ \ Scale_{ij,t} + \beta_8 Industry\ Size_{ij,t} + \beta_9 Same\ SIC2_{ij,t} + \boldsymbol{Year}\\
&\quad +\ \boldsymbol{\varepsilon_{ijt}} \qquad (4)
\end{aligned}
$$

Relatively higher overtime variation in the industry portfolio of small audit firms allows me to examine my hypotheses in a change setting. I examine how across-industry relatedness influences which new industry-pairs are more likely to be added to the small audit firms' existing industry portfolios. I use the following OLS model to implement the change analysis about small audit firms' industry specialty portfolio change.

Table 4: Small Audit Firms' Industry Portfolio Choice

Panel A: What Industry Pairs Small Audit Firms Jointly Operate In

VARIABLES	Mode1 1	Model 2	Model 3
Across Industry Language Similarity	0.161***	0.168***	0.154***
	(6.568)	(6.608)	(6.991)
Fraction Industries Between Pair	0.032***	0.049***	0.058***
	(2.604)	(3.844)	(4.384)
Within Industry Language Similarity	-0.042***	-0.037***	-0.036***
	(-19.92)	(-16.810)	(-16.570)
Vertical Relatedness	-0.151***	-0.151***	-0.137***
	(-10.28)	(-10.530)	(-8.087)
Industry Stability	-0.007*	-0.007*	-0.007*
	(-1.840)	(-1.686)	(-1.793)
Pair Likelihood if Random	0.603***	0.593***	0.575***
	(11.910)	(11.690)	(11.780)
Economies of Scale	-0.006***	-0.006**	-0.006**
	(-2.801)	(-2.447)	(-2.445)
Industry Size	0.040***	0.043***	0.045***
	(22.55)	(24.430)	(26.590)
Same SIC2	0.002	0.003	0.003*
	(1.262)	(1.628)	(1.653)
Constant	0.004*	-0.000	0.000
	(1.920)	(-0.384)	(0.140)
Observations	22,334	23,319	23,907
Adjusted R-squared	0.453	0.445	0.437
Year Fixed	YES	YES	YES

This table reports the coefficients and associated t-statistics (in parentheses) for the following model:

$$
\begin{aligned}
&Fraction\ Audit\ Firms\ in\ Existing\ Industry\ Pair_{ijt} \\
&\quad = \beta_0 + \beta_1 Average\ Across\ Industry\ Language\ Similarity_{ij,t} \\
&\quad + \beta_2 Fraction\ Industries\ Between\ Pair_{ij,t} \\
&\quad + \beta_3\ With\ Indusry\ Language\ Similarity_{ij,t} + \ \beta_4\ Vertical\_Relatedness_{ij,t} \\
&\quad + \beta_5 Industry\ Stability_{ij,t} + \beta_6 Pair\ Likelihood\ if\ Random_{ij,t} \\
&\quad + \beta_7 Economies\ of\ Scale_{ij,t} + \beta_8 Industry\ Size_{ij,t} + \beta_9 Same\ SIC2_{ij,t} + \boldsymbol{Year} \\
&\quad + \ \boldsymbol{\varepsilon_{ijt}} \qquad (4)
\end{aligned}
$$

One observation is one SIC3 industry pair in a year derived from the set of all permutations of feasible pairings of industries in which at least one small audit firm operates in. The dependent variable is the fraction of small audit firms having auditing practice in the given SIC3 industry pair (relative to the total number of small audit firms). In model 1, I restrict my small audit firm sample to audit firms with between 5 and 60 clients. In model 2, I restrict my small audit firm sample to audit firms with between 5 and 70 clients and in model 3 I restrict my small audit firm sample to audit firms with between 5 and 80 clients. For detailed variable definitions, please refer to Appendix A. Statistical significance level of the estimated coefficients are indicated as *** p<0.01, ** p<0.05, * p<0.1 using two-tailed tests.

Table 4 Continued

Panel B: Newly Added SIC3 Industries Pairs

VARIABLES	Model 1	Model 2	Model 3
Across Industry Language Similarity	0.870***	0.902***	0.829***
	(3.301)	(3.447)	(3.373)
Fraction Industries Between Pair	0.870***	0.934***	1.099***
	(4.152)	(4.528)	(5.386)
Within Industry Language Similarity	-1.339***	-1.253***	-1.233***
	(-23.850)	(-23.050)	(-23.340)
Vertical Relatedness	-1.970***	-1.976***	-1.652***
	(-6.309)	(-6.565)	(-4.710)
Industry Stability	-0.057***	-0.056***	-0.060***
	(-3.978)	(-3.994)	(-4.363)
Pair Likelihood if Random	4.075***	3.423***	3.065***
	(9.103)	(8.569)	(8.428)
Economies of Scale	-0.083*	-0.077*	-0.065
	(-1.844)	(-1.745)	(-1.534)
Industry Size	0.866***	0.880***	0.889***
	(41.730)	(45.440)	(49.050)
Same SIC2	0.040	0.0421	0.040
	(1.385)	(1.444)	(1.422)
Constant	0.646***	0.575***	0.507***
	(12.340)	(11.260)	(10.400)
Observations	22,725	24,436	24,330
Adjusted R-squared	0.261	0.251	0.246
Year Fixed	YES	YES	YES

This table reports the results of the following model:

$$\begin{aligned}
&Fraction\ Audit\ Firms\ in\ Industry\ Pair\ Added_{ij,t} \\
&\quad = \beta_0 + \beta_1 Across\ Industry\ Language\ Similarity_{ij,t} \\
&\quad + \beta_2 Fraction\ Industries\ Between\ Pair_{ij,t} \\
&\quad + \beta_3\ With\ Indusry\ Language\ Similarity_{ij,t} + \ \beta_4\ Vertical_Relatedness_{ij,t} \\
&\quad + \beta_5 Industry\ Stability_{ij,t} + \beta_6 Pair\ Likelihood\ if\ Random_{ij,t} \\
&\quad + \beta_7 Economies\ of\ Scale_{ij,t} + \beta_8 Industry\ Size_{ij,t} + \beta_9 Same\ SIC2_{ij,t} + \boldsymbol{Year} \\
&\quad + \ \boldsymbol{\varepsilon_{ijt}} \qquad\qquad (5)
\end{aligned}$$

One observation is one pair of SIC3 industries in a year derived from the set of all permutations of feasible pairings of industries in which at least one small audit firm operates in. The dependent variable is the fraction of small audit firms having auditing practice in a given SIC3 industry-pair that it did not have during previous year (relative to the total number of small audit firms that did not have auditing practice in such SIC3 industry-pair during the previous year), multiplied by 100 for convenience. In model 1, I restrict my sample to audit firms with between 5 and 60 clients. In model 2, I restrict my sample to audit firms with between 5 and 70 clients and in model 3 I restrict my sample to audit firms with between 5 and 80 clients. For detailed variable definitions please refer to Appendix A. Statistical significance level of the estimated coefficients are indicated as *** p<0.01, ** p<0.05, * p<0.1 using two-tailed tests.

$$
\begin{aligned}
&Fraction\ Audit\ Firms\ in\ Industry\ Pair\ Added_{ij,t} \\
&\quad = \beta_0 + \beta_1 Across\ Industry\ Language\ Similarity_{ij,t} \\
&\quad + \beta_2 Fraction\ Industries\ Between\ Pair_{ij,t} \\
&\quad + \beta_3\ With\ Indusry\ Language\ Similarity_{ij,t} + \ \beta_4\ Vertical\ Relatedness_{ij,t} \\
&\quad + \beta_5 Industry\ Stability_{ij,t} + \beta_6 Pair\ Likelihood\ if\ Random_{ij,t} \\
&\quad + \beta_7 Economies\ of\ Scale_{ij,t} + \beta_8 Industry\ Size_{ij,t} + \beta_9 Same\ SIC2_{ij,t} + \boldsymbol{Year} \\
&\quad + \ \boldsymbol{\varepsilon_{ijt}} \qquad (5)
\end{aligned}
$$

Where *Fraction Audit Firms in Industry Pair Added* is defined as the fraction of small audit firms that add a new industry pair in a given fiscal year (relative to the total number of small audit firms that did not have auditing practice in such SIC3 industry-pair in previous year). The ratio is multiplied by 100 for convenience. For both equation (4) and (5), please refer to section 3.3 and Appendix A for detailed definition of variables.

Table 4 Panel A reports the estimation results of model (4). The results indicate that small audit firms are more likely to jointly operate in industry-pairs with high across industry language similarity and with more between-industries. This is consistent with what we already observed with the industry specialty portfolio decisions of Big 4 audit firms. However, unlike Big 4 audit firms, small audit firms are less likely to jointly operate in industry-pairs with high within industry language similarity (i.e., more specialized industries). This is consistent with the view that the limited capacity of small audit firms impedes them from taking advantage of the economies of scale of highly specialized industries. Given the observation that Big 4 audit firms are more likely to specialize in more specialized industries and enjoy cost efficiency (Bills et al. 2014), the inherent disadvantages of small audit firms in terms of cost efficiency would further discourage them from specializing in industries with high within industry language similarity.

Table 4 Panel B reports the estimation results of model (5). It shows that small audit firms are more likely to add new industry-pairs with high across industry language similarity, more between-industries, and when within industry language similarity is low. Such results further confirm the findings provided in *Table 4 Panel A.*

6.4 Arthur Andersen Event Analysis

The analysis in section 4.2 shows that the industry specialty portfolio decisions of Big 4 audit firms depend not only on the industries' standalone characteristics but also on across industry relatedness in the product space. In this section, I examine whether synergies arising from across industry relatedness in the product space affect the auditor-client contracting features. I use the collapse of Arthur Andersen in 2002 as an exogenous supply shock in the audit market (the demise of the Arthur Anderson caused the transition from Big 5 to Big 4) and test whether the average across-industry language similarity between the industries of former Arthur Andersen clients and the industry specialty portfolio of Big 4 in 2001 associates with the final matching between former Arthur Andersen clients and Big 4 audit firms in 2003.[23] I hypothesize that former Arthur Andersen clients are more likely to be matched with Big 4 audit firm with an industry specialty portfolio that is more aligned with them. The Arthur Andersen event is a good setting to test my conjecture because the event is likely to be exogenous to the audit contract termination between Arthur Andersen and its clients except for Enron. Gerakos and Syverson (2015) rely on the exogenous nature of the audit market supply shift caused by the collapse of Arthur Andersen to estimate the price elasticity of client firms' demand. They provide ample evidence to support the exogenous nature of contract termination to existing auditor-client characteristics.

For each former Arthur Andersen client, I calculate the *Average Across Industry Language Similarity* (*AAILS*) between its industry and Big 4 audit firms' industry specialty portfolios.

[23] Note that in my main analysis the sample period is from 2003 to 2013.

Table 5: Matching Between Former Arthur Andersen Clients and the Big 4 Audit Firms

Panel A: Ratio of Correct Prediction

Column 1	Column 2		Column 3		Column 4		Column 5
Audit Firms	Firms In Big 4 Specialized Industries		Firms In Big 4 Specialized Industries		Firms Not in Big 4 Specialized Industries		Matched Audit Firm In 2003
	SIC3 Industry Specialization		Average Across Industry Similarity		Average Across Industry Similarity		
	2001	2003	2001	2003	2001	2003	
		28		26		26	PwC
		66		57		27	E&Y
PwC	185	44	160	38	88	14	Deloitte
		47		39		21	KPMG
		9		6		21	PwC
		21		21		49	E&Y
E&Y	76	22	69	18	127	31	Deloitte
		24		24		26	KPMG
		17		22		13	PwC
		13		20		17	E&Y
Deloitte	87	45	113	53	65	21	Deloitte
		12		18		14	KPMG
		4		4		19	PwC
		10		12		23	E&Y
KPMG	26	6	31	8	103	23	Deloitte
		6		8		38	KPMG
Total	(374)		(374)		(383)		
Ratio of Correct Prediction		27%		29%		35%	

This table shows the matching results between former Arthur Andersen clients (in 2001) and Big 4 audit firms in 2003. There are in total 755 (374 + 383) former Arthur Anderson clients that were eventually matched with Big 4 audit firms in 2003. Among the 755 former Arthur Andersen clients, 374 firms (subsample 1) were in industries in which Big 4 were the specialized audit firms in 2001 and 383 (subsample 2) firms were not in industries in which Big 4 were specialized. In column 2 and column 3, I sort subsample 1 according to which Big 4 audit firm was the firm's industry specialized auditor firm. In column 4, I sort subsample 2 according to which Big 4 audit firm had industry specialty portfolio most aligned with the firm in 2001. For example, cell row 1 column 2 reports that there were 185 former Arthur Andersen clients in industries where PwC was the specialized audit firm in 2001, and in 2003, among these former Arthur Andersen clients, 28 firms were matched with PwC, 66 firms were matched with E&Y, 44 firms were matched with Deloitte, and 47 firms were matched with KPMG.

Table 5: Continued

Panel B: Mean Test

	Asset Based Industry Specialization						Sale Based Industry Specialization					
Sample	Whole sample		Subsample 1		Subsample 2		Whole sample		Sample 1		Sample 2	
Mean Test	2.34***	2.41**	2.40*	2.42	2.27***	2.39*	2.39***	2.39***	2.43	2.40*	2.36**	2.39**
	(-3.96)	(-2.33)	(-1.65)	(-1.35)	(-3.88)	(-1.94)	(-2.69)	(-2.72)	(-1.34)	(-1.82)	(-2.44)	(-2.03)
Observations	757	750	374	374	383	376	757	757	383	383	374	374
	All Pairs	Top 3 Pairs	All Pairs	Top3 Pairs	All Pairs	Top3 Pairs	All Pairs	Top3 Pairs	All Pairs	Top3 Pairs	All Pairs	Top3 Pairs

This table reports the one sample t-test results between the average rank of the matching results and 2.5. I calculate the average across industry language similarity between former Arthur Andersen clients' industries and the Big 4 audit firms' industry specialty portfolios and rank Big 4 audit firms according to the average across industry language similarity. The Big 4 audit firm with the highest average across industry language similarity score has rank 1 and the Big 4 audit firm with the lowest average across industry language similarity has rank 4. If the matching is independent of the potential synergies between the firm's industry and Big 4 audit firms' industry specialty portfolios, the average rank of the matching results should be 2.5 (i.e., (1+2+3+4)/4). In the left panel, Big 4's industry specialty is identified according to the market share of their clients' total assets. In the right panel, Big 4's industry specialty is identified according to the market share of their clients' total sales. There are in total 755 (374 + 383) former Arthur Anderson clients that were eventually matched with Big 4 audit firms in 2003 (whole sample). Among the 755 former Arthur Andersen clients, 374 firms (subsample 1) were in industries in which Big 4 were the specialized audit firms in 2001 and 383 (subsample 2) firms were not in industries in which Big 4 were specialized.

$$Average\ Across\ Industry\ Language\ Similarity_{\mathrm{ij}} = \frac{\sum_{\mathrm{k=1}}^{\mathrm{J}} AILS_{\mathrm{ijk}}}{K} \qquad (2)$$

Where the average across industry language similarity between former Arthur Andersen client i and Big 4 audit firm j is the average across industry language similarity between the firm's industry and each of the Big 4 audit firm j's specialized industry k. Then I rank the Big 4 audit firms according to the calculated *Average Across Industry Language Similarity*. The Big 4 audit firm with the highest *AAILS* is ranked 1 and the Big 4 audit firm with the lowest *AAILS* is ranked 4. My conjecture is that former Arthur Andersen clients are more likely to be matched with the Big 4 audit firm with lower rank (i.e., higher average across industry language similarity).

Table 5 Panel A shows the matching results between the former Arthur Andersen clients and the Big 4 audit firms. There were 757 former Arthur Andersen clients in 2001 that eventually matched with Big 4 audit firms in 2003. Among the 757 former Arthur Andersen clients, 374 firms are in industries in which Big 4 audit firms were the specialized audit firms in 2001. I first check, in column 1, whether former Arthur Andersen clients are simply more likely to be matched with the audit firm that is the specialized audit firm in the firm's industry. Cell column 1 row 2 reports that 185 former Arthur Andersen client firms had PwC as their specialized auditor in 2001; however, only 28 of them were still matched with PwC in 2003. In contrast, there were 87 former Arthur Andersen clients had Deloitte as their specialized auditor in 2001; and 45 of them matched with Deloitte in 2003. The overall ratio of correct prediction for the 374 former Arthur Andersen clients is 27%. In column 2 and column 3, I check whether former Arthur Andersen clients are more likely to be matched with Big 4 audit firm with industry specialty portfolio that is most aligned with the firm (i.e., highest *AAILS*). To make comparison with column 1, column 2 shows the matching results for the same sample used in column 1. The overall ratio of correct prediction is 29%, not statistically different from that in column 1. In column 4, the sample is the 383 former Arthur Andersen clients that did not have a specialized Big 4 audit firm in 2001. For this sample, the ratio of correct prediction is 35%, which is a 30% increase compared to column 1 and the difference is statistically significant. From this simple analysis, I cannot reach a clear conclusion on which matching mechanism is more important. In the following, I formally test whether former Arthur Andersen clients are more likely to be matched with audit firm whose industry specialty portfolio is more aligned with them.

If the matching is independent of the potential synergies between a firm's industry and Big 4 audit firms' industry specialty portfolios, the average rank of the matched audit firm should be 2.5 (i.e., (1+2+3+4)/4). On the other hand, if former Arthur Anderson clients are more likely to be matched with Big 4 audit firm with industry specialty portfolio that is most aligned with the firm the average rank of the matched audit firm should be less than 2.5. In *Table 5 Panel B,* I report the results of t-test on whether the average rank of the matching is lower than 2.5. In the left panel, I use the market share of the total assets of client firms to identify Big 4 audit firms' industry specialty portfolios. In the right panel, I use the market share of the total sale of client firms to identify Big 4 audit firms' industry specialty portfolios. The average rank of the matching is statistically lower than 2.5. Consistent with the basic statistics in *Table 5 Panel A*, such effect is more pronounced for former Arthur Andersen clients that did not have a specialized Big 4 audit firms in 2001. One potential explanation could be that the fact that those former Arthur Andersen clients, for which Big 4 audit firms were their specialized audit firms at 2001, were not matched with Big 4 audit firms at 2001 imply that there were other important factors affecting the matching between these firms and the Big 4 audit firms such that they were not matched with the specialized audit firms at the first place (at year 2001).

Bills et al. (2014) document evidence that the cost efficiency associated with more homogeneous industry could be passed to client firms in terms of lower audit fee. Would the across-industry synergies also be reflected in auditor-client contracting process such as audit fee? Two non-exclusive mechanisms exist to support the conjecture that former Arthur Andersen clients with better match (higher Average Across-industry Language Similarity) would enjoy lower audit fee. First, audit firm whose industry specialty portfolio is more aligned with the firm may have an initial cost advantage and can afford to offer lower audit fees. Second, consistent with the "low-balling" theory (Chan, 1999) it is also possible that audit firms offer lower audit fee in order to capture future potential synergistic benefits. *Table 6* reports the OLS regression results of audit fee discounts on the *Average Across Industry Language Similarity*. I control firm size, ROA, leverage, inventory and receivable, number of segments, and firms' foreign sale ratio.[24] Audit fee discount is the difference between predicted audit fees for the firm and the realized audit fees.[25] For robustness, the test is conducted by using firms' audit fee in 2003 and 2004 respectively. The results show that former Arthur Andersen clients enjoyed more audit fee discounts when they were matched with audit firms whose industry specialty portfolios had higher synergies with the firm's industry (i.e., higher *Average Across Industry Language Similarity).*

[24] Please refer to Appendix A for detailed variable definition. The reason that I only control for some generic firm characteristics and do not control for more event-specific characteristics such as whether the contract termination is initiated by the audit firm or the clients is that in the Arthur Anderson setting such contract termination is mostly likely exogenous.

[25] I follow Gerakos and Syverson (2015) to use Random Forest method to estimate firms' predicted audit fee.

Table 6: Audit Fees Discount: Former Arthur Andersen Clients

	Model 1		Model 2	
Variables	2003	2004	2003	2004
Average Across Industry Language Similarity	-6.154**	-4.698	-5.765**	-6.200*
	(-2.35)	(-1.471)	(-2.206)	(-1.771)
Size	0.044**	0.088***	0.042**	0.088***
	(2.184)	(3.864)	(2.117)	(3.932)
ROA	-0.103	-0.240*	-0.087	-0.255**
	(-0.720)	(-1.947)	(-0.622)	(-2.076)
Leverage	0.010	0.020	0.028	0.050
	(0.078)	(0.18)	(0.217)	(0.453)
Inventory and Receivables	0.244	0.282*	0.250*	0.321**
	(1.605)	(1.738)	(1.673)	(2.031)
Ln(Segments)	0.073	0.053	0.075	0.046
	(1.300)	(0.855)	(1.336)	(0.746)
Foreign Sale Ratio	0.112	0.084	0.127	0.095
	(0.736)	(0.498)	(0.838)	(0.559)
Constant	-0.288**	-0.505***	-0.294***	-0.497***
	(-2.574)	(-3.816)	(-2.676)	(-3.859)
Observations	555	488	555	488
Auditor Fixed	YES	YES	YES	YES
Adjusted R-squared	0.022	0.075	0.021	0.079

This table reports the results of OLS regression of audit fee discount on average across industry language similarity between former Arthur Andersen clients and the Big 4 audit firms' industry specialty portfolios. T-statistics are based on robust standard errors.[26] The dependent variable is former Arthur Andersen clients' audit fee discount (i.e., realized audit fees minus the predicted audit fees). The variable of interest is the average across industry language similarity between former Arthur Andersen client's industry and the Big 4 audit firms' industry specialty portfolios. In model 1, Big 4's industry specialty is identified according to the market share of their clients' total assets. In model 2, Big 4's industry specialty is identified according to the market share of their clients' total sales. Statistical significance level of the estimated coefficients are indicated as *** $p<0.01$, ** $p<0.05$, * $p<0.1$ using two-tailed tests.

[26] Clustering with relatively few groups (4 audit firms in my case) could provide noisy estimates of standard errors. Clustering by auditor give similar and even tighter estimates.

7. Additional Analysis

7.1 Text-Based Industry Classification

While I used textual information about firms' product description to construct my across-industry relatedness measures, my empirical analyses are based on three-digit SIC industry classification. Industries could also be defined based on the textual information firms use to describe their products. It is shown that text-based industry classification could better explain across-industry variation (Hoberg and Philips, 2015). In this section, I replicate my analyses of the industry portfolio choice of small audit firms by using the text-based industry classification provided by Hoberg and Philips (2015). In particular, I use their Fixed Industry Classification with 300 Industries (FIC-300), which is a set of text-based industries chosen to be roughly as granular as the SIC3 industry classification. Across industry relatedness measures are constructed similarly as for the SIC3 industry classification. I do not include this as my primary analyses due to the fact that many variables are not available using text-based industry classifications. Textual information needed to construct text-based industries only became available starting in 1996 by SEC requirement. I do not use the text-based industry classification for the analyses of the Big 4 audit firms' industry portfolio choices analysis. This is because the text-based industry classification only includes firms traded on major exchanges and the sample loss caused by this constraint makes it hard to identify Big 4 audit firms' industry specialty.

Table 7 reports the results of model (4) and model (5) but with the text-based Fixed Industry Classification with 300 industries. Note that the controlling variable *Same SIC2* and *Vertical Relatedness* are not included since they are not available when using the text-based industry classification. The results are consistent with what are reported in *Table 6*. Small audit firms are more likely to operate in industry-pairs with high across industry language similarity and with more between industries. Estimated coefficients on *Within Industry Language Similarity* are not significant in *Table 7 Panel A*.

Table 7: Small Audit Firms' Industry Portfolio Choice: Using Text Based Industry Classification

Panel A: What FIC300 Industry Pairs Small Audit Firms Jointly Operate In

Variables	Model 1	Model 2	Model 3
Across Industry Language Similarity	0.160***	0.204***	0.209***
	(9.380)	(6.474)	(5.958)
Fraction Industries Between Pair	0.040***	0.045***	0.050***
	(6.619)	(6.178)	(5.612)
Within Industry Language Similarity	-0.001	0.002	-0.000
	(-0.705)	(1.312)	(-0.285)
Industry Stability	-0.005***	-0.002**	-0.004***
	(-4.268)	(-2.292)	(-3.072)
Pair Likelihood if Random	0.504***	0.574***	0.749***
	(12.050)	(15.280)	(15.520)
Economies of Scale	-0.001	0.0015	-0.003*
	(-0.758)	(1.008)	(-1.885)
Industry Size	0.012***	0.013***	0.021***
	(10.880)	(13.860)	(16.280)
Constant	-0.007***	-0.008***	-0.008***
	(-4.213)	(-4.049)	(-3.908)
Observations	31,525	33,696	36,764
Adjusted R-squared	0.452	0.445	0.450
Year Fixed	YES	YES	YES

This table reports the results of model (4) using FIC300 text-based industry classification instead of SIC3 industry classification. One observation is one pair of FIC300 industries in a year derived from the set of all permutations of feasible pairings of FIC300 industries in which at least one small audit firm operates in. The dependent variable is the fraction of small audit firms having auditing practice in the given FIC300 industry pair (relative to the total number of small audit firms). In model 1, I restrict my small audit firm sample to audit firms with between 5 and 60 clients. In model 2, I restrict my small audit firm sample to audit firms with between 5 and 70 clients and in model 3 I restrict my small audit firm sample to audit firms with between 5 and 80 clients. For detailed variable definitions please refer to Appendix A. Statistical significance level of the estimated coefficients are indicated as *** p<0.01, ** p<0.05, * p<0.1 using two-tailed tests.

Table 7 Continued

Panel B: Newly Added FIC300 Industries Pairs

VARIABLES	Model 1	Model 2	Model 3
Across Industry Language Similarity	0.210***	0.236***	0.218***
	(7.048)	(7.719)	(6.878)
Fraction Industries Between Pair	0.094***	0.112***	0.079***
	(4.928)	(5.780)	(4.082)
Within Industry Language Similarity	-0.003	-0.010**	-0.026***
	(-0.724)	(-2.100)	(-5.779)
Industry Stability	0.004**	0.006***	0.002
	(1.971)	(2.864)	(1.230)
Pair Likelihood if Random	0.096***	0.111***	0.135***
	(3.738)	(3.814)	(3.993)
Economies of Scale	-0.033***	-0.036***	-0.045***
	(-6.737)	(-7.534)	(-9.359)
Industry Size	0.062***	0.066***	0.085***
	(27.520)	(30.140)	(35.490)
Constant	0.023***	0.041***	0.049***
	(4.067)	(7.346)	(8.698)
Observations	31,525	33,696	36,764
Adjusted R-squared	0.060	0.068	0.077
Year Fixed	YES	YES	YES

This table reports the results of model (5) when using FIC300 text-based industry classification instead of SIC3 industry classification. One observation is one pair of FIC300 industries in a year derived from the set of all permutations of feasible pairings of industries in which at least one small audit firm operates in. The dependent variable is the fraction of small audit firms having auditing practice in a given FIC300 industry-pair that it did not have during previous year (relative to the total number of small audit firms that did not have auditing practice in such FIC300 industry-pair during the previous year), multiplied by 100 for convenience. In model 1, I restrict my sample to audit firms with between 5 and 60 clients. In model 2, I restrict my sample to audit firms with between 5 and 70 clients and in model 3 I restrict my sample to audit firms with between 5 and 80 clients. For detailed variable definitions please refer to Appendix A. Statistical significance level of the estimated coefficients are indicated as *** p<0.01, ** p<0.05, * p<0.1 using two-tailed tests.

7.2 Geographical Industry Clusters

Small audit firms often are geographically constrained and thus geographical industry clustering could have a direct effect on small audit firms' industry portfolio composition. For instance, industry-pairs derived from the same industrial cluster would be more likely to be observed in small audit firms' industry portfolio. To mitigate this concern, I rerun model 4 and 5 controlling for whether an SIC3 industry-pair is derived from the same industrial clusters or not. *Same Industry Cluster* is a dummy variable which takes value of one if two SIC3 industries belong to the same U.S. industrial cluster and takes value of zero otherwise. I use the industrial clusters defined by Delgado et al. (2014) to construct the dummy variable *Same Industry Cluster*.[27]

Table 8 reports the results of model 4 and model 5 with additional controlling variable *Same Industry Cluster*. The results do show that small audit firms are more likely jointly operate in industry-pairs within the same industrial cluster. However, controlling for the effect of geographical industrial clustering does not change the main results.

[27] Please refer to their website for more information on their definition of industrial clusters http://clustermapping.us/

Table 8: Small Audit Firms' Industry Portfolio Choice: Controlling for Industry Clusters

Panel A: What Industry Pairs Small Audit Firms Operate In

VARIABLES	Mode1 1	Model 2	Model 3
Across Industry Language Similarity	0.147***	0.154***	0.138***
	(5.824)	(5.893)	(6.141)
Fraction Industries Between Pair	0.037***	0.054***	0.063***
	(3.032)	(4.258)	(4.796)
Within Industry Language Similarity	-0.038***	-0.033***	-0.032***
	(-18.760)	(-15.660)	(-15.330)
Vertical Relatedness	-0.168***	-0.168***	-0.154***
	(-10.910)	(-11.190)	(-8.387)
Same Industry Cluster	0.008***	0.008***	0.008***
	(8.458)	(8.256)	(8.756)
Industry Stability	-0.007*	-0.007*	-0.007*
	(-1.904)	(-1.722)	(-1.865)
Pair Likelihood if Random	0.600***	0.591***	0.572***
	(11.850)	(11.640)	(11.730)
Economies of Scale	-0.006***	-0.006**	-0.006**
	(-2.727)	(-2.417)	(-2.398)
Industry Size	0.0395***	0.0421***	0.044***
	(22.380)	(24.290)	(26.430)
Same SIC2	-0.002	-0.001	-0.001
	(-1.003)	(-0.544)	(-0.613)
Constant	0.002	-0.003	-0.002
	(1.069)	(-1.184)	(-0.692)
Observations	22,334	23,319	23,907
Adjusted R-squared	0.456	0.448	0.440
Year Fixed	YES	YES	YES

This table reports the results of model (4) with control of the effect of industry clusters. One observation is one pair of SIC3 industries in a year derived from the set of all permutations of feasible pairings of industries in which at least one small audit firm operates. The dependent variable is the fraction of small audit firms having auditing practice in the given SIC3 industry pair (relative to the total number of small audit firms). In model 1, I restrict my small audit firm sample to audit firms with between 5 and 60 clients. In model 2, I restrict my small audit firm sample to audit firms with between 5 and 70 clients. And in model 3 I restrict my small audit firm sample to audit firms with between 5 and 80 clients. For detailed variable definition, please refer to Appendix A. Statistical significance level of the estimated coefficients are indicated as *** $p<0.01$, ** $p<0.05$, * $p<0.1$ using two-tailed tests.

Table 8 Continued

Panel B: Newly Added SIC3 Industries Pair

VARIABLES	Model 1	Model 2	Model 3
Across Industry Language Similarity	0.150***	0.158***	0.143***
	(6.131)	(6.190)	(6.545)
Fraction Industries Between Pair	0.040***	0.058***	0.067***
	(3.274)	(4.506)	(5.074)
Within Industry Language Similarity	-0.038***	-0.033***	-0.032***
	(-18.640)	(-15.500)	(-15.170)
Vertical Relatedness	-0.173***	-0.173***	-0.159***
	(-11.140)	(-11.430)	(-8.569)
Same Industry Cluster	0.010***	0.010***	0.011***
	(9.082)	(8.895)	(9.300)
Industry Stability	-0.008**	-0.007*	-0.008**
	(-2.056)	(-1.870)	(-1.992)
Pair Likelihood if Random	0.597***	0.588***	0.569***
	(11.810)	(11.600)	(11.690)
Economies of Scale	-0.007***	-0.007***	-0.007***
	(-3.195)	(-2.862)	(-2.875)
Industry Size	0.039***	0.041***	0.043***
	(22.240)	(24.190)	(26.340)
Same SIC2	-0.003*	-0.002	-0.002
	(-1.756)	(-1.246)	(-1.342)
Constant	0.003	-0.002	-0.000
	(1.430)	(-0.849)	(-0.347)
Observations	22,334	23,319	23,907
Adjusted R-squared	0.457	0.449	0.441
Year Fixed	YES	YES	YES

This table reports the results of model (5) with control of the effect of industry clusters. One observation is one pair of SIC3 industries in a year derived from the set of all permutations of feasible pairings of industries in which at least one small audit firm operates in. The dependent variable is the fraction of small audit firms having auditing practice in a given SIC3 industry-pair that it did not have during previous year (relative to the total number of small audit firms that did not have auditing practice in such SIC3 industry-pair during the previous year), multiplied by 100 for convenience. In model 1, I restrict my sample to audit firms with between 5 and 60 clients. In model 2, I restrict my sample to audit firms with between 5 and 70 clients and in model 3 I restrict my sample to audit firms with between 5 and 80 clients. For detailed variable definitions please refer to Appendix A. Statistical significance level of the estimated coefficients are indicated as *** p<0.01, ** p<0.05, * p<0.1 using two-tailed tests.

8. Conclusion

The audit market can be viewed as a network of industries and the structure of this network is affected by audit firms' industry portfolio decisions. Despite the long standing economic theory on firms' strategic product provision decisions (Hotelling, 1929, Lancaster, 1966, 1979), extant research in the auditing literature often takes audit firms' industry specialization decisions as given and we know little about how audit firms make their industry portfolio choice. In this study, I examine how the linkages between industries in the product space, which reflect the trade-off between specialization and coordination and potential synergistic value between industries (Becker and Murphy, 1992, Cremer et al., 2007), affect audit firms' industry portfolio choice.

In this study, I use text-based product space measures to capture across-industry linkages. My findings show that both Big 4 and small audit firms are more likely to specialize in industry-pairs that have more similar product language and a higher number of between-industries in the product space, consistent with audit firms choosing industries in the network that allow them to capitalize on existing investments and economize on coordination costs. These findings also suggest that audit firms' industry portfolio decisions are endogenous—audit firms strategically take into account the potential synergistic benefits and coordination costs arising from the across-industry linkages in the product space. Further, using the collapse of Arthur Andersen as an exogenous supply shock in the audit market, I find consistent results that across-industry linkages also affect the auditor-client contracting process.

A distinctive feature of my study is that I consider the audit market as a network of industries in the product space and explore how audit firms' industry portfolio choice affects the formation of this network. Future studies could explore other potential factors such as audit partners' social ties (Guan et al., 2015, Horton et al., 2015) and audit firms' organizational form (e.g., Lennox et al., 2012, Firth et al., 2012). Understanding what factors give rise to the observed audit market outcomes is crucial for design and evaluation of policy interventions because it is the underlying market structure that determines the market outcome (Demsetz, 1973).

Appendices

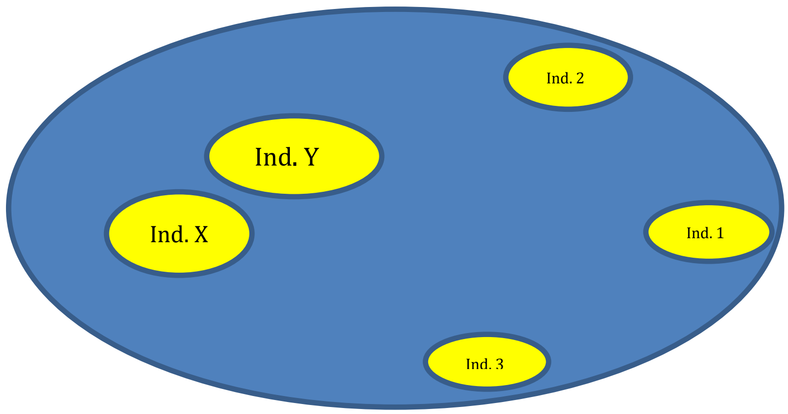
Appendix A: Variable Definition

Variable	Definition
Analysis on the product space location of Big4 audit firms' industry specialty	
Industry Pair	Binary dummy variable which takes value of one if the SIC3 industry-pair is derived from the same Big 4 audit firm's industry specialty portfolio.
Across Industry Language Similarity	The across-industry language similarity measure is the average textual cosine similarity of all permutation of a firm-pair where one firm is in SIC3 industry i and the other is from SIC3 industry j.
Within Industry Language Similarity	The average of the within industry product language similarity of the two SIC3 industry in a pair. (Within industry similarity is the average pairwise similarity of firms in a given industry.)
Fraction Industries Between Pair	Number of SIC3 industries locating between an industry-pair; a third industry is between two industries in a given industry-pair if the third industry is closer in textual distance to each industry in the pair than the two industries in the pair are to each other.
Vertical Relatedness	Vertical relatedness between two SIC3 industries according to the Bureau of Economic Analysis's Use Table of Benchmark Input-Output Accounts of the US Economy. I follow th methodology described by Fan and Goyal (2006) to define the vertical relatedness of two SIC3 industries.
Pair Likelihood if Random	The probability of a randomly draw industry-pair where one industry is from SIC3 industry i and one is from a SIC3 industry j. It is calculated as (# of firms in SIC3 industry i / total number of firms in the market) * (# of firms in SIC3 industry j / total number of firms in the market).
Economy of Scale	The average economy of scale of an SIC3 industry pair. The economy of scale is based on the estimation of a Cobb-Douglas production function over ten years, with sales being the dependent variable.
Industry Stability	The average of the absolute values of the logarithmic change in the number of firms in the given two SIC3 industries in an industry-pair over the previous year.
Industry Size	Decile rank of the average number of firms of the two SIC3 industries in an industry-pair.
Analysis on the product space location of small audit firms	
Fraction Audit Firms in Existing Industry Pair	The fraction of small audit firms having audit practice in a SIC3 industry-pair (relative to the total number of small audit firms).
Fraction Audit Firms in Industry Pair Added	The fraction of small audit firms having auditing practice in a given SIC3 industry-pair that it did not have such industry-pair during previous year (relative to the total number of small audit firms that did not have auditing practice in such SIC3 industry-pair during previous year); multiplied by 100 for convenience.
Arthur Andersen Event Analysis and Audit Fee Predication	
Across Industry Language Similarity	The average of product space distance between prior Arthur Andersen client's industry and Big 4 audit firm's industry specialty portfolio.
Size	The natural logarithm of the firm's total assets.
ROA	Income before extraordinary items deflated by total assets.
Leverage	The ratio of total debt (short- and long-term debt) to total assets of the firm.
Inventory and Receivables	Sum of inventory and receivable, divided by beginning total assets.
Ln(Segments)	The natural logarithm of number of a firm's industrial segments at SIC3 level.
Foreign Sale Ratio	Foreign sale is the percentage of the firm's sale generated outside of United States.

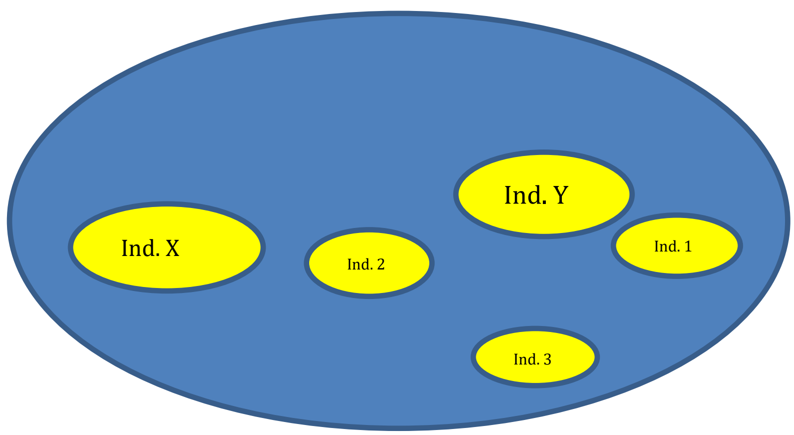
Appendix B: Industry Characteristics in the Product Space

Figures 1A to 1C depict the concepts of the three measures of industry characteristics in the product space, respectively, in a simplified two-dimensional space. For a given industry, the actual dimension of the industry's product space will be *N* where *N* is the total number of non-trivial words used in the industry's product descriptions. Figure 1A depicts the concept of across-industry similarity where industry X and industry Y are closer to each other than other industry-pairs. Figure 1B shows the concept of between-industries. It shows industry 2 is between Industry X and Industry Y. Figure 1C shows the concept of the within-industry similarity where industry X and industry Y are more spread-out in the product space than ind.1 ind.2 and ind.3.

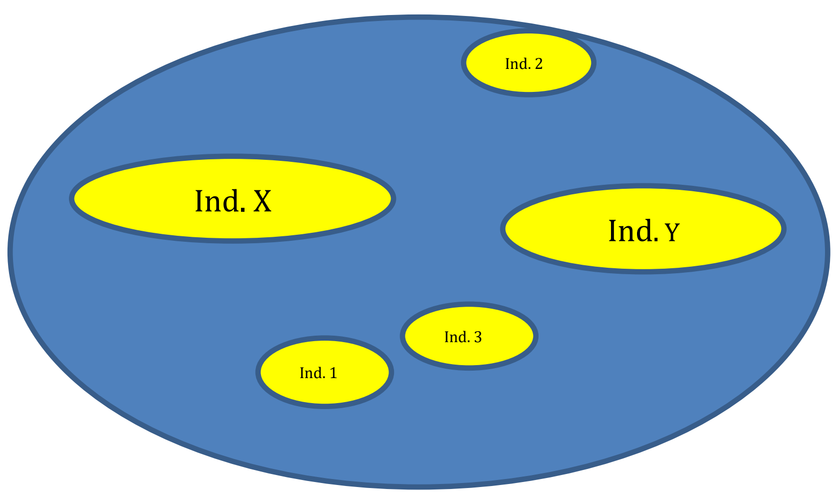
1 A



1 B

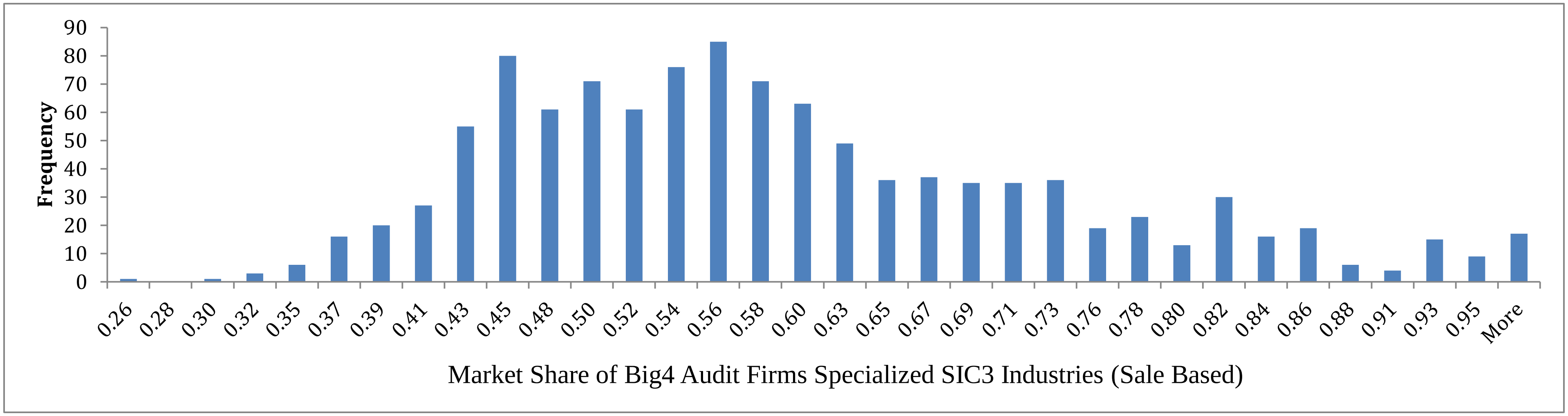


1 C



Appendix C: Descriptive Statistics of Big 4 Audit Firms' Industry Specialization (Based on Client Firms' Sale)

Panel A: Distribution of SIC3 Industry Market Share



Panel B: Distribution of Big 4' Specialized SIC3 Industries

Auditor	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
PwC	33	34	32	37	30	31	30	30	31	27	30	34	29	408
EY	25	32	35	33	30	29	27	25	30	27	27	26	24	370
Deloitte	15	24	22	24	31	33	31	27	22	26	23	25	22	325
KPMG	13	16	16	15	13	13	13	12	13	14	13	17	17	185
AA	19	0	0	0	0	0	0	0	0	0	0	0	0	19
Total	105	106	105	109	104	106	101	94	96	94	93	102	92	1,307

Panel A shows the distribution of Big 4 audit firms' market share in their specialized SIC3 industries. I define one audit firm as a specialized audit firm in one SIC3 industry if the audit firm is the market leader in the industry and has at least 10% lead of market share (sale based) to the second ranked audit firm. The market share is based on the client firms' total sale. Panel B shows the distribution of specialized SIC3 industries over the Big-N firms (PwC: PricewaterhouseCoopers, EY: Ernst & Young, Deloitte: Deloitte Touche Tohmastsu Limited, KPMG, and AA: Arthur Andersen) and over the period 2001 – 2013.

References

Ahern, K. R., and J. Harford. 2014. The importance of industry links in merger waves. *The Journal of Finance* 69 (2): 527–576.

Amihud, Y., and B. Lev. 1981. Risk Reduction as a Managerial Motive for Conglomerate Mergers. *The Bell Journal of Economics* 12 (2): 605–617.

Becker, G. S., and K. M. Murphy. 1992. The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics* 107 (4): 1137.

Bell, T.; F. Marrs; I. Solomon; and H. Thomas. Auditing 0rganizations Through a Strategic Systems Lens: The KPMG Business Measurement Process. Montvale, N.J.: KPMG Peat Marwick LLP, 1997.

Bills, K. L., D. C. Jeter, and S. E. Stein. 2014. Auditor industry specialization and evidence of cost efficiencies in homogenous industries. *The Accounting Review*.

Bolton, P., and M. Dewatripont. 1994. The firm as a communication network. *The Quarterly Journal of Economics*: 809–839.

Brown, S. V, and W. R. Knechel. 2013. *Auditor-Client Compatibility and Audit Firm Selection*. Working paper, March 12, 2013.

Cahan, S. F., J. M. Godfrey, J. Hamilton, and D. C. Jeter. 2008. Auditor specialization, auditor dominance, and audit fees: The role of investment opportunities. *The Accounting Review* 83 (6): 1393–1423.

Cairney, T. D., and G. R. Young. 2006. Homogenous industries and auditor specialization: An indication of production economies. *Auditing: A Journal of Practice & Theory* 25 (1): 49–67.

Carter, J. R. 1977. In search of synergy: A structure-performance test. *The Review of Economics and Statistics*: 279–289.

Chan. D., G. Feltham, and D. Simunic. 1997. A spatial analysis of competition in the market for audit services. Working paper. University of British Columbia.

Chan, D. K. 1999. "Low‐Balling" and Efficiency in a Two‐Period Specialization Model of Auditing Competition. *Contemporary Accounting Research* 16 (4): 609–642.

Craswell, A. T., J. R. Francis, and S. L. Taylor. 1995. Auditor brand name reputations and industry specializations. *Journal of accounting and economics* 20 (3): 297–322.

Cremer, J., L. Garicano, and A. Prat. 2007. Language and the Theory of the Firm. *The Quarterly Journal of Economics*: 373–407.

Danos, P., and J. W. Eichenseher. 1982. Audit industry dynamics: Factors affecting changes in client-industry market shares. *Journal of Accounting Research*: 604–616.

———. 1986. Long-term trends toward seller concentration in the US audit market. *Accounting Review*: 633–650.

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

Delgado, M., M. E. Porter, and S. Stern. 2014. Clusters, convergence, and economic performance. *Research Policy* 43 (10): 1785–1799.

Demsetz, H. 1973. Industry structure, market rivalry, and public policy. *Journal of Law and economics*: 1–9.

Eichenseher, J. W., and P. Danos. 1981. The analysis of industry-specific auditor concentration: Towards an explanatory model. *Accounting Review*: 479–492.

Emerson, J. 1993. KPMG Peat Marwick: Setting the new practice framework standard. *Professional Services Review* 1: 1–4.

Fan, J. P. H., and V. K. Goyal. 2006. On the Patterns and Wealth Effects of Vertical Mergers*. *The Journal of Business* 79 (2): 877–902.

Ferguson, A., J. R. Francis, and D. J. Stokes. 2003. The effects of firm-wide and office-level industry expertise on audit pricing. *The accounting review* 78 (2): 429–448.

Ferguson, C., M. Pinnuck, and D. J. Skinner. 2014. The evolution of audit market structure and the emergence of the Big 4: Evidence from Australia. *Chicago Booth Research Paper* (14-13).

Firth, M., P. L. L. Mo, and R. M. K. Wong. 2012. Auditors' Organizational Form, Legal Liability, and Reporting Conservatism: Evidence from China. Contemporary Accounting Research 29 (1): 57–93.

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

Francis, J. R., and D. T. Simon. 1987. A test of audit pricing in the small-client segment of the US audit market. *Accounting Review*: 145–157.

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.

Garicano, L., and T. N. Hubbard. 2009. Specialization, firms, and markets: The division of labor within and between law firms. *Journal of Law, Economics, and Organization* 25 (2): 339–371.

Gerakos, J., and C. Syverson. 2015. Competition in the audit market: Policy implications. *Journal of Accounting Research*.

Gilson, R. J., and R. H. Mnookin. 1985. Sharing among the human capitalists: An economic inquiry into the corporate law firm and how partners split profits. *Stanford law review*: 313–392.

Gong, Q., O. Z. Li, Y. Lin, and L. Wu. 2014. On the Benefits of Audit Market Consolidation: Evidence from Merged Audit Firms. The Accounting Review.

Gramling, A. A., and D. N. Stone. 2001. Audit firm industry expertise: A review and synthesis of the archival literature. *Journal of accounting literature* 20: 1.

Guan, Y., L. N. Su, D. Wu, and Z. Yang. 2015. Do school ties between auditors and client executives influence audit outcomes? Journal of Accounting and Economics.

Gul, F. A., S. Y. K. Fung, and B. Jaggi. 2009. Earnings quality: Some evidence on the role of auditor tenure and auditors' industry expertise. *Journal of Accounting and Economics* 47 (3): 265–287.

Hart, O., and J. Moore. 2005. On the Design of Hierarchies: Coordination versus Specialization. *Journal of Political Economy* 113 (4): 675–702.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23 (10): 3773–3811.

———. 2015. Conglomerate Industry Choice and Product Differentiation. *Working Paper*.

Hoberg, G., and G. M. Phillips. 2015. Text-Based Network Industries and Endogenous Product Differentiation. *conditionally accepted Journal of Political Economy*.

Hogan, C. E., and D. C. Jeter. 1999. Industry specialization by auditors. *Auditing: A Journal of Practice & Theory* 18 (1): 1–17.

Hope, O. K., and J. C. Langli. 2010. Auditor independence in a private firm and low litigation risk setting. *The Accounting Review* 85 (2): 573–605.

Horton, J., T. Irem, and W. Anthony. 2015. Audit Partner Performance: A Network Perspective. Working Paper

Hotelling, H. 1929. Stability in Competition. *Economic Journal* 39: 41–57.

Johnson, W. B., and T. Lys. 1990. The market for audit services: Evidence from voluntary auditor changes. *Journal of accounting and economics* 12 (1): 281–308.

Johnstone, K. M., and J. C. Bedard. 2003. Risk management in client acceptance decisions. *The Accounting Review* 78 (4): 1003–1025.

———. 2004. Audit firm portfolio management decisions. *Journal of Accounting Research* 42 (4): 659–690.

Lancaster, K. J. 1966. A new approach to consumer theory. *The journal of political economy*: 132–157.

Lancaster, K. J. 1979. Monopolistic competition with a non-uniform distribution of consumers, The Economics Workshops, Columbia University Discussion Paper 15

Lennox, C., and B. Li. 2012. The consequences of protecting audit partners' personal assets from the threat of liability. Journal of Accounting and Economics 54 (2-3): 154–173.

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.

Mazzeo, M. J. 2002. Product choice and oligopoly market structure. *RAND Journal of Economics*: 221–242.

Minutti-Meza, M. 2013. Does auditor industry specialization improve audit quality? *Journal of Accounting Research* 51 (4): 779–817.

Numan, W., and M. Willekens. 2012. An empirical test of spatial competition in the audit market. *Journal of Accounting and Economics* 53 (1–2): 450–465.

Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM computing surveys (CSUR)* 34 (1): 1–47.

Shu, S. Z. 2000. Auditor resignations: Clientele effects and legal liability. *Journal of Accounting and Economics* 29 (2): 173–205.

Simunic, D. A. 1980. The pricing of audit services: Theory and evidence. *Journal of accounting research*: 161–190.

Simunic, D. A., and M. T. Stein. 1990. Audit risk in a client portfolio context. Contemporary Accounting Research 6 (2): 329–343.