Discretionary Disclosure in Financial Reporting: An Examination Comparing Internal Firm Data to Externally Reported Segment Data — Source link

Daniel A. Bens, Philip G. Berger, Steven J. Monahan, Steven J. Monahan

Institutions: INSEAD, University of Chicago, University of Utah

Published on: 01 Sep 2009 - Social Science Research Network

Topics: Agency cost, Agency (sociology), Strategic business unit and Discretion

DISCRETIONARY DISCLOSURE IN FINANCIAL REPORTING:

AN EXAMINATION COMPARING INTERNAL FIRM DATA TO EXTERNALLY REPORTED SEGMENT DATA

by

Daniel A. Bens *
University of Arizona

Philip G. Berger *
University of Chicago

and

Steven J. Monahan *
INSEAD

CES 09-28 September, 2009

Abstract

We use confidential, U.S. Census Bureau, plant-level data to investigate aggregation in external reporting. We compare firms' plant-level data to their published segment reports, conducting our tests by grouping a firm's plants that share the same four-digit SIC code into a "pseudo-segment." We then determine whether that pseudo-segment is disclosed as an external segment, or whether it is subsumed into a different business unit for external reporting purposes. We find pseudo-segments are more likely to be aggregated within a line-of-business segment when the agency and proprietary costs of separately reporting the pseudo-segment are higher and when firm and pseudo-segment characteristics allow for more discretion in the application of segment reporting rules. For firms reporting multiple external segments, aggregation of pseudo-segments is driven by both agency and proprietary costs. However, for firms reporting a single external segment, we find no evidence of an agency cost motive for aggregation.

Keywords: manufacturing plants; micro-level data; segment reporting; discretionary disclosure; agency costs; proprietary costs

Data Availability: All data are available from public sources, except for the Census micro-level data.

* The research in this paper was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Chicago Census Research Data Center. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Lynn Riggs and Frank Limehouse of the Chicago RDC provided excellent advice and guidance about the project. We gratefully acknowledge support for this research that was provided by the National Science Foundation (NSF awards no. SES-0004335 and ITR-0427889). Bens gratefully acknowledges financial support provided by the Ratoff Family Fellowship at the University of Arizona. Berger gratefully acknowledges financial support provided by the University of Chicago Booth School of Business and its Neubauer Family Faculty Fellows program. Monahan gratefully acknowledges financial support provided by the INSEAD Alumni Fund. We have benefitted from the comments of two anonymous reviewers, Rebecca Hann, Ole-Kristian Hope, Mary Stanford, and Wayne Thomas (the editor) as well as workshop participants at the 2008 Tel Aviv Conference in Accounting, the U.S. Census Bureau Center for Economic Studies, the 2007 UC-Davis Conference on Financial Markets Research, and workshops at the following universities: Arizona, British Columbia, Colorado, Connecticut, INSEAD, Maryland, Michigan, Penn State, Texas Christian, Toronto, Virginia (Darden), and Wisconsin.

1. INTRODUCTION

We use a unique database provided by the Census Bureau that contains *confidential*, plant-level data to investigate two questions. First, what factors motivate managers of publicly-traded firms to conceal information when making discretionary disclosure decisions? Second, how do private competitors affect both the disclosure decisions themselves and industry-level statistics used to construct explanatory variables?

Having access to the Census data gives us three advantages over prior research. First, we are better able to observe management's private information endowment.[1] Past segment disclosure research (e.g., Harris 1998; Botosan and Stanford 2005; Berger and Hann 2007) has examined managers' aggregation decisions by evaluating the amount of segment-level information in *publicly* disseminated financial statements. A fundamental limitation of this approach is that, by definition, the researcher does not consider the underlying source data management observes when making the disclosure decision. However, because the Census data contain confidential information on a firms' operations that are less aggregated than the information presented in public filings, we are able to: (1) observe internal activities conducted by the firm even if these activities are not separately disclosed in the segment footnotes and (2) more accurately measure variables that we use to test our hypotheses.

Our ability to observe management's private information endowment is made more valuable by the inextricable linkage between the agency cost and proprietary cost hypotheses for nondisclosure that we investigate. Agency costs of segment disclosure may arise when disaggregated segment data provide information about a firm's diversification strategy indicative of unresolved agency problems. If agency costs were the only plausible motive for nondisclosure, there would be no point in withholding

[1] The use of the term "private information endowment" is justified by the great lengths to which Census goes to preserve the confidentiality of its micro-level data and the privacy of its survey respondents. Census micro-level data are only accessible by individuals that have Special Sworn Status, which is only granted after a number of background checks and other evaluations are performed. Individuals with Special Sworn Status must agree to preserve the confidentiality of the data and failure to do so leaves them open to criminal penalties under Title 13, U.S.C. In particular, wrongful disclosure of confidential data is punishable by a fine of as much as $250,000 and/or imprisonment for up to five years. Moreover, while statistics derived from Census micro-level data may be publicly disseminated (e.g., in research papers) these statistics are first subject to a rigorous review by Census' Disclosure Review Board. This review is done to ensure that the respondents' privacy is maintained and that no confidential micro-level data are released.

information from rational market participants as they would infer that nondisclosure signals the worst possible outcome (Grossman 1981; Milgrom 1981). Thus, the plausibility of the proprietary cost motive is necessary for the agency cost motive to potentially exist. Moreover, the agency and proprietary cost motives are reasonable only if outsiders cannot use *publicly available* disclosures to fully unravel these motives for aggregation. Taken together these observations imply: (1) the agency and proprietary cost motives should be studied simultaneously and (2) data from *publicly disseminated* financial statements have limitations for developing unambiguous empirical inferences about either of these motives for nondisclosure. Hence, our access to the confidential Census data at the disaggregated operating level is a key advantage for addressing our hypotheses and it allows us to overcome some limitations of prior research.

The second advantage of using the Census data is that the Census surveys all (i.e., both publicly-traded and privately-held) U.S. establishments in a particular industry. Hence, we are able to more accurately measure industry-level phenomena that potentially affect managers' disclosure decisions. In addition, we are able to develop a new proprietary cost measure that equals the proportion of total industry sales made by private firms. We use this variable to test whether public firms competing in industries with high concentrations of private firms are more likely to mimic the non-disclosure policies of their private competitors.

The final advantage arising from our access to the Census data is that we are able to examine the extent to which agency costs and proprietary costs affect the segment disclosures of firms that report only one segment in their external financial statements (i.e., single-segment firms). Because prior researchers were unable to observe financial information at a level more disaggregated than a financial statement segment, they typically suppressed single-segment firms and focused on the disclosure choices of firms that report multiple segments in their external financial statements (i.e., multisegment firms). By analyzing single-segment firms we provide some initial evidence on whether (and to what extent) the disclosure choices of these firms differ from those of multisegment firms.

As discussed above, we simultaneously evaluate the proprietary cost and agency cost motives for nondisclosure. Prior studies of segment reporting (e.g., Harris 1998; Botosan and Stanford 2005) have primarily focused on the proprietary cost motive, which posits that nondisclosure occurs in order to conceal proprietary information of value to competitors, suppliers, or regulators. Although the proprietary cost motive has received considerable attention, there is not a clear consensus regarding its descriptive validity.

There are three main issues with the prior empirical evidence regarding the proprietary cost motive. First, the evidence is conflicting. For example, consider the evidence regarding the role of industry concentration, which is a commonly-used measure of product market competition. Bamber and Cheon (1998) find that whether a firm *provides* earnings forecasts is negatively related with industry concentration; however, Verrecchia and Weber (2006) find that whether a firm asks the Securities and Exchange Commission (SEC) to *withhold* proprietary information from its filings is also negatively related with industry concentration. Hence, the former study's findings are consistent with *less* informative disclosure in more concentrated industries whereas the results of the latter study are consistent with *more* informative disclosure in more concentrated industries.

Second, there are concerns about the way proprietary cost proxies have been measured. Again, consider product market competition proxies. Typically these proxies are calculated using Compustat data, which include publicly-traded (but generally not privately-held) firms. However, Ali, Klasa, and Yeung (2009) show that industry concentration measures calculated using only Compustat data are weak proxies for total (public and private firm) industry concentration. Ali et al. show that using U.S. Census industry concentration measures often reverses prior results in the literature (including research that examines the segment reporting decision) that are based on Compustat concentration metrics.

Finally, Berger and Hann (2007) point out that it is difficult to disentangle the proprietary and agency cost hypotheses from the extant evidence. They argue that much of the prior evidence consistent with the proprietary cost hypothesis is also consistent with an alternative agency cost hypothesis that posits disclosures are withheld as a result of conflicts of interest between managers and shareholders. Under the agency cost hypothesis segment nondisclosure results from managers attempting to reduce the potential costs to themselves from segment disclosures that reveal underperformance. Berger and Hann's findings are consistent with the agency cost hypothesis, but are at best mixed with respect to the proprietary cost hypothesis.

In light of the above, we develop constructs for both proprietary costs and agency costs so that we can examine the relative strength of these respective phenomena after controlling for several "non-strategic" determinants of disclosure (e.g., financial accounting standards and capital market phenomena). We test our hypotheses using a sample of 1,625 firm-years from 1987, 1992, and 1997, with coverage in both Compustat and the Longitudinal Research Database (LRD), maintained by the Center for Economic Studies at the Bureau of the Census. As discussed further in Section 3.1, the Census' LRD surveys virtually the entire population of U.S. manufacturing plants every five years. Ours is the first paper within the accounting literature to utilize these plant-level data. Our objective is to assess how managers choose to aggregate internal data into external reports.

Our main analyses are conducted using logistic regressions at the "pseudo-segment" level. We aggregate all of the firm's LRD plants within the same four-digit Standard Industrial Classification (SIC) code together into one pseudo-segment. The dependent variable is an indicator set equal to one for pseudo-segments with four-digit SIC codes that match either the primary or secondary SIC code of a disclosed segment for that firm on the Compustat database ("disclosed" pseudo-segments) and zero otherwise ("hidden" pseudo-segments).

While we analyze both single- and multisegment firms our primary focus is on the aggregation decisions made by managers of multisegment firms. We focus on multisegment firms for three reasons. First, as discussed above, previous studies on segment reporting tend to forgo analyses of single-segment firms. Second, we argue that *a priori* multisegment firms have greater motive, means and opportunity to aggregate. Finally, we provide empirical evidence that is consistent with the previous argument. Specifically, our empirical results suggest that managers of multisegment firms behave more strategically when making aggregation decisions than do managers of single-segment firms.

Our most robust results within the multisegment sample are that the likelihood a pseudo-segment is disclosed separately is negatively related to: (1) whether the pseudo-segment receives inefficient transfers of funds from the remainder of the firm and (2) the speed of abnormal profit adjustment exhibited by firms in the pseudo-segment's industry. The interpretation of the relation between inefficient transfers and the likelihood that a pseudo-segment is separately disclosed is consistent with the agency cost motive: managers suppress information about internal capital transfers that are not in the best interests of shareholders. The interpretation of the relation between the likelihood of disclosure and the industry speed of abnormal profit adjustment is less clear-cut, however. This result is similar to the finding documented in Harris (1998), who interprets it as a manifestation of the proprietary cost motive. Nonetheless, as discussed in Berger and Hann (2007), this result may also be interpreted as consistent with the agency cost motive.

To understand better the relation between the likelihood of disclosure and industry speed of profit adjustment we separate our multisegment sample into observations that have value-reducing diversification programs (i.e., negative excess value firms) and observations that have value-enhancing diversification programs (i.e., nonnegative excess value firms) and then we re-estimate our model separately for each subsample.[2] We show that the industry speed of abnormal profit adjustment is negatively related to the probability of pseudo-segment disclosure in the nonnegative excess value subsample but not in the negative excess value subsample. Given that the nonnegative excess value subsample is less likely to contain firms with agency cost motives for pseudo-segment aggregation, these results support the proprietary cost interpretation discussed in Harris (1998).

Within the multisegment sample we also document that the likelihood of a pseudo-segment being disclosed is positively associated with the pseudo-segment's industry-adjusted profitability and negatively associated with a labor power proxy that equals the ratio of aggregate industry wages to aggregate

[2] We consider a multisegment firm's diversification program to be value-reducing (value-enhancing) if the excess value measure developed in Berger and Ofek (1995) is less than zero (nonnegative). The excess value measure represents an estimate of diversification's effect on firm value. In particular, we compare the sum of the imputed stand-alone values of a firm's individual segments to the firm's actual value. Negative excess values indicate a value loss from combining the segments into a single firm whereas positive excess values indicate a value gain.

industry sales. The first result is consistent with the agency cost motive for nondisclosure as it implies that information about less profitable operations is hidden. The second result is consistent with the proprietary cost motive as it implies that firms act in their shareholders' interest by withholding information from other rent-seeking stakeholders. While both of these results are manifest in our main set of tests, they are not robust to some of our sensitivity tests. In contrast, the results related to the industry speed of abnormal profit adjustment and to inefficient transfers are robust.

As discussed above, we also analyze the single-segment firms in our sample. These analyses suggest that managers of single-segment firms behave less strategically than managers of multisegment firms when making disclosure decisions. Moreover, when managers of single-segment firms are strategic in their disclosure decisions they do so in order to avoid revealing information to their private competitors. In particular, for our single-segment sample only one of our agency/proprietary cost proxies is associated with the likelihood of disclosure. This proxy equals the proportion of industry sales attributable to privately-held firms. Hence, a single-segment firm is less likely to identify a pseudo-segment as being in the firm's main or secondary industry if that pseudo-segment operates in an industry with a relatively high concentration of privately-held firms. This novel result offers a potentially new insight about the nature of proprietary costs as it suggests that publicly-traded firms seek to minimize disclosure levels when they compete with privately-held firms that have no legal duty to publicly disclose.

The rest of the paper is organized as follows. Section 2 discusses related literature and our hypotheses. Section 3 presents our sample selection, data description, and research design. Our empirical results are discussed in sections 4 and 5, with our concluding comments offered in section 6.

2. HYPOTHESIS DEVELOPMENT AND BENEFITS OF USING CENSUS DATA

2.1. Hypotheses

Whether or not a business unit is reported as a separate segment falls under the umbrella of the general choice of whether or not to disclose, which is analyzed by Verrecchia (1983). In particular, as modeled by Hayes and Lundholm (1996), the segment disclosure decision relates to *fineness* or *disaggregation* of the data. That is, a consolidated business unit's performance and financial position is required to be included within the aggregate information reported by the entity in its consolidated financial statements. However, management may report the unit separately (the unit is disaggregated) or it may not (the unit is aggregated). Based on models such as those in Verrecchia (1983) and Hayes and Lundholm (1996), prior empirical studies have explored whether proprietary information costs reduce disclosure of industry segments via aggregation.

The extant evidence regarding the proprietary cost motive for aggregation is decidedly mixed. Consistent with this motive, Harris (1998) finds that operations in less competitive industries are less likely to be reported as separate industry segments; and, Botosan and Stanford (2005) provide evidence that managers hide profitable segments operating in less competitive industries. However, Ali et al. (2009) show that Harris's finding is attributable to the use of Compustat data, which generally exclude privately-held firms, to measure industry competition.[3] Specifically, when Ali et al. use U.S. Census-based industry concentration measures, they find no association between the level of industry competition and the decision to separately report operations in that industry. Furthermore, Botosan and Harris (2000) find no association between proprietary costs and the decision to voluntarily increase segment disclosure frequency. Finally, Berger and Hann (2007) generally fail to find results consistent with the hypothesis that proprietary costs are an important motive for withholding line-of-business segments. Hence, the relation between proprietary costs and aggregation decisions is an open empirical issue and, thus, our first hypothesis is that there is a negative association between proprietary costs and the likelihood that a particular pseudo-segment will be reported as a separate external segment.

Managers may also face agency costs of segment disclosure, which arise when disaggregated segment data provide information that is indicative of unresolved agency problems. Segment data provide information about a company's diversification strategy and its transfers of resources across divisions. Prior research finds evidence consistent with internal capital markets in conglomerates transferring funds

[3] Ettredge, Kwon, Smith and Stone (2006) also use industry-based proprietary cost measures calculated via Compustat. They conclude that multisegment firms tend to smooth profits across segments when proprietary costs captured by their industry-based measures are high. Ali et al. do not examine the sensitivity of these results.

across segments in a suboptimal manner (Berger and Ofek 1995; Lamont 1997; Shin and Stulz 1998; Rajan, Servaes and Zingales 2000). Several studies indicate that diversified firms trade at a discount relative to stand-alone firms (Lang and Stulz 1994; Berger and Ofek 1995) and that the diversification discount is associated with measures of agency problems (Dennis, Dennis, and Sarin 1997; Berger and Ofek 1999).

Berger and Hann (2003) find that firms moving to more disaggregated segment reporting under the mandated change of SFAS 131 experienced an increased diversification discount, implying that managers concealed information about agency problems under SFAS 14.[4] The same implication arises from Bens and Monahan (2004), who find that greater voluntary segment disaggregation is associated with a smaller diversification discount. Berger and Hann (2007) test this implication and find evidence consistent with the withholding of segment data being motivated by the desire to conceal agency problems. Following these studies, our second hypothesis therefore predicts that managers face agency cost motives to withhold segment data via aggregation.

2.2. Benefits of Using Census Data

Our access to confidential Census data at the disaggregated operating level allows us to overcome some limitations of prior research. For example, Harris (1998) collects the number of SIC codes identified by *Standard & Poor's* reviews of annual reports and SEC filings. She then checks whether each three-digit SIC code identified by *S&P* is listed as a primary or secondary code for an externally reported segment. This approach cannot capture the *magnitude* of the activity in the SIC code. If the magnitude of operations within that code is small, Harris's proxy for expected segment disclosure is measured with error as it treats the small activity as one that requires an individual segment disclosure. We overcome this drawback because we observe the amount of sales (as well as contribution margin and investment activity) for each plant and thus for each pseudo-segment.

[4] SFAS 14 is FASB Statement No. 14, *Financial Reporting for Segment of a Business Enterprise* (FASB 1976). SFAS 131 is FASB Statement No. 131, *Disclosures about Segments of an Enterprise and Related Information* (FASB 1997).

Botosan and Stanford (2005) and Berger and Hann (2007) attempt to overcome some of the limitations of observing only ex-post disclosed data by exploiting the change in segment disclosure rules from SFAS 14 to SFAS 131. These papers compare the degree of information aggregation originally reported under the old standard to the restatement of those reports under the new standard. However, these studies are limited by examining only the restated year(s) of reporting under the old standard and by the fact that the new standard is still subject to considerable managerial discretion. For example, practitioners claimed that SFAS 131 was not adopted uniformly by all firms, and that segment disclosure was still manipulated by some companies (Sanders, Alexander and Clark 1999). Our paper is the first to examine how the confidential, internally reported, line-of-business data influence an external reporting disclosure decision.

3. SAMPLE SELECTION, DATA AND METHODOLOGY

3.1. Sample Selection & Data

Our sample selection begins with all annual firm observations from the years 1987, 1992 and 1997 for which the firm is listed on Compustat's Annual Industrial, Research, or Full Coverage files and also covered within the Longitudinal Research Database (LRD), maintained by the Center for Economic Studies at the Bureau of the Census. The LRD is made up of two databases. The first is the *Census of Manufactures* (*CM*), which is conducted every five years. The unit of analysis is a manufacturing plant, which the Census refers to as an *establishment*. Extremely small establishments are excluded from the *CM*, although the Census does impute estimates for them based on Internal Revenue Service and Social Security Administration data; all other establishments are required by law to respond truthfully to the *CM* (U. S. Code Title 13, § 224). The second LRD database is the *Annual Survey of Manufactures* (*ASM*), which is conducted in all non-*CM* years. Through the end of our sample period, all establishments with more than 250 employees as of the most recent *CM* are included in the *ASM* panel.[5]

[5] Our sample period ends in 1997. In 1999, the employee certainty criterion was raised from 250 to 500 employees and in 2004 it was raised from 500 to 1,000 employees. Smaller establishments are selected with a probability

We use only the *CM* years to ensure that the LRD covers not only all of the establishments owned by our sample firms, but also nearly all of the establishments in the industry. Because many of our independent variables are industry-based measures, this research design choice decreases measurement error. It also has the advantage of making our research more comparable to studies that use *publicly available* industry measures from the *CM* that are obtainable from the Census's web site. Although such measures are limited to a very few industry-level variables such as the four firm concentration index and the Herfindahl concentration index, recent studies such as Ali, Klasa and Yeung (2009) and Tang (2009) demonstrate that the incorporation of private firms into the industry-level measures can be quite useful in certain contexts.[6]

The sample begins in 1987 because that is the first *CM* year for which we have access to Compustat's segment data and for which Compustat reports segment SIC codes. The sample ends in 1997 because that is the last year in which firms reported segments under the SFAS 14 rules and also the last year in which the Bureau of the Census classified industries using SIC codes (NAICS codes are now used). Ending the sample with the completion of the SFAS 14 regime allows us to restrict our attention to *industry-based* segment reporting as opposed to the *internal-management* segment reporting perspective of SFAS 131. Nevertheless, our findings are likely to be quite applicable to the SFAS 131 regime as it continues to allow managers some discretion in aggregating plant-level data into reportable segments and the vast majority of the internal-management segments under SFAS 131 turn out to be based on either industry or geography. More importantly, our main goal is not to explain the use of discretion in financial reporting aggregation only in the context of segment reporting, but rather to use segment reporting to examine discretion in financial reporting aggregation more generally. The somewhat greater discretion that seems to have been available to managers under SFAS 14 is thus a desirable feature for our research.

---

proportional to their size. Once selected, the establishment remains in the *ASM* for the four years following the *CM*. Comprehensive descriptions of the LRD may be found in McGuckin and Pascoe (1988) or at the web site of the Center for Economic Studies (http://www.ces.census.gov).

[6] The most recent online data from the *CM* are available at http://www.census.gov/econ/census07, where links are also available to censuses from 2002, 1997 and 1992.

We aggregate all of the firm's establishments that operate within the same four-digit SIC code together into one "pseudo-segment." We require our sample firms to have establishments that collectively operate in at least two pseudo-segments, thus ensuring that all sample firms have some potential for aggregation of pseudo-segments into financial reporting segments.

Because the LRD covers only U.S. manufacturing establishments, we require that the firm be domiciled in the U.S. and that its primary SIC code on Compustat fall within the manufacturing sector (i.e., between 2000 and 3999). Even so, many firms have a portion of their production that occurs either outside the U.S. or outside of manufacturing. Moreover, while Compustat correctly eliminates intersegment sales within a firm some double-counting of interplant sales does occur in the Census data. Therefore, the final step in our sample selection process is to eliminate observations where there is a poor match between the firm's sales per the LRD and its sales per Compustat. In particular, we remove observations where the ratio of the firm's LRD *total value of shipments* (*TVS*) to Compustat sales is less than 0.75 or greater than 1.25. *TVS* is the Census term for sales.[7] The ratio of *TVS* to sales will be less than 0.75 if the firm has a great deal of non-manufacturing sales and/or it has production facilities outside the U.S. that generate sales. The ratio exceeds 1.25 if intersegment sales have not been eliminated properly by Census.

3.2. Dependent Variable

Following Harris (1998), we evaluate a dichotomous dependent variable, *MATCH*, that measures whether a pseudo-segment is disclosed as a separate external segment per Compustat. In particular, *MATCH* equals one if the pseudo-segment's four-digit SIC code per the Census matches either the primary or secondary SIC codes of one of the firm's line-of-business segments per Compustat, and zero otherwise. Under SFAS 14, enterprises were required to classify line-of-business segment information using the *industry approach*. Thus, we assess the extent to which a four-digit SIC-based classification of

[7] *TVS* is defined as the "selling value f.o.b. plant after discounts and allowances and excluding freight charges" (*Value of Product Shipments: 2005 – Annual Survey of Manufactures*, Appendix A, issued November 2006).

industries for all the firm's internally tracked data match the externally reported industry segments chosen by management.[8]

Segment disclosure, and especially which firm operations constitute a separable unit for reporting purposes, has been a contentious issue for decades and it has provided fertile ground for researchers interested in disclosure. Researchers' interest derives from the fact that while SFAS 14 required disclosure about firms' operations in different industries (paragraph 1), the FASB explicitly stated that no single industry classification system (such as the SIC code) "is, by itself, suitable to determine industry segments for purposes of this Statement" (paragraph 12). Paragraph 12 continues with the following:

> [N]o single set of characteristics is universally applicable in determining the industry segments of all enterprises, nor is any single characteristic determinative in all cases. Consequently, determination of an enterprise's industry segments must depend to a considerable extent on the management of the enterprise.

This discretion afforded to management is what makes segment reporting an interesting disclosure choice. However, this discretion also makes it difficult for researchers to determine which industries *should* have been disclosed by management. In our setting, we rely on further guidance from SFAS 14 to generate predictions of unbiased disclosure. Specifically, paragraph 13 states:

> An enterprise's existing profit centers – the smallest units of activity for which revenue and expense information is accumulated for internal planning purposes – represent a logical starting point for determining the enterprise's industry segments.

The definition of *profit center* coincides with Census's choice to collect plant-level data. This, in turn, justifies our maintained assumption that a pseudo-segment (i.e., a group of profit centers sharing the same four-digit SIC code) is the starting point from which management uses its discretion under SFAS 14 when deciding whether to aggregate the pseudo-segment with others or to report it separately in the segment footnotes. We are not asserting that each pseudo-segment *should* be reported separately; rather, our maintained assumption is that each pseudo-segment *has the potential* to be reported separately.

[8] We are interested only in the line-of-business (LOB) segment reporting choice for two reasons. First, as a practical matter, the LRD data only cover U.S. manufacturing plants and the extent of geographical diversification in these data is thus limited to within-U.S. variation. Second, with the exception of Hope and Thomas (2008) who focus on geographic diversification, the extant literature has focused on agency problems within the context of line-of-business diversification.

Given the FASB explicitly states that four-digit SIC codes are not a sufficient basis for *all* firms to determine their reportable segments, our aggregation of plants by common four-digit SIC code into pseudo-segments is admittedly ad hoc. On the other hand, the FASB provides no preferable alternative. Moreover, defining industries with SIC codes has a long history in the accounting, economics and finance literatures, with Bhojraj, Lee and Oler (2003) noting that SIC codes have been available since 1939 and that more than 90% of studies that use a general industry classification scheme use SIC codes despite their shortcomings. Thus, we believe it reasonable to rely on this classification scheme. Nonetheless, we acknowledge that the use of four-digit SIC codes may introduce measurement error. Hence, in section 5.3.1 we: (1) report a validity check of our disclosure measure, *MATCH*, using a comparison based on the data set collected by Berger and Hann (2003; 2007) and (2) discuss a sensitivity check in which we define industries on the basis of three-digit SIC codes.

3.3. Model of Managers' Segment Reporting Choice

We examine managers' segment reporting decisions by estimating the following logit regression at the pseudo-segment level (pseudo-segment, firm and year subscripts are omitted):

$$\begin{aligned} MATCH = {} & \beta_0 + \beta_1 \times I\_PROFIT + \beta_2 \times PROFITADJ + \beta_3 \times PRIVATE + \beta_4 \times TRANSIN \\ & + \beta_5 \times LOENT\_HISUB + \beta_6 \times LABOR + \beta_7 \times FSIZE + \beta_8 \times FOURFIRM \\ & + \beta_9 \times RELSIZE + \beta_{10} \times SEGDIVERSITY + \beta_{11} \times NUMEST \\ & + \beta_{12} \times CEN\_CMPSTAT + \beta_{13} \times INDMATCHRATE + \beta_{14} \times NEC + \varepsilon \end{aligned} \tag{1}$$

As discussed above, the dependent variable, *MATCH*, is a dichotomous variable with the value of one if the pseudo-segment's four-digit SIC code per the Census matches a primary or secondary SIC code of a Compustat line-of-business segment reported by the firm, and zero otherwise.

3.3.1. Treatment Variables

I_PROFIT, our measure of abnormal profits, is a ratio. The numerator of *I_PROFIT* equals the pseudo-segment's gross margin percentage less the average gross margin percentage for all establishments in the pseudo-segment's industry.[9] This number is then divided by the standard deviation of gross margin percentage for all establishments in the pseudo-segment's industry. Gross margin percentage is calculated using Census data and equals (*TVS* – *CM* – *SW*) / *TVS*. In this formula *TVS* is total value of shipments, *CM* is cost of materials, and *SW* is salaries and wages. We acknowledge that this variable does not fully capture *total* pseudo-segment profitability (i.e., it ignores some expenses and, more importantly, asset turnover). By industry-adjusting, however, we reduce the likelihood of our profit measure having a low correlation with a total profitability measure such as return on assets (ROA) because the negative correlation typically observed between total asset turnover and gross margin is smaller within industries than across industries. We do not attempt to calculate an ROA-based measure because the Census data do not contain sufficient information about establishment-level assets.

I_PROFIT is a direct measure of pseudo-segment profits that captures how well a pseudo-segment performs relative to its industry. It is pseudo-segment profits rather than publicly available firm-level (or external segment-level) profits that managers may try to hide via aggregation. Thus, because it is based on confidential information that is not observable by outsiders, *I_PROFIT* is a superior proxy to publicly-observable measures of profit. The coefficient relating *I_PROFIT* to disclosure will be positive if the agency cost hypothesis dominates and only strongly performing units tend to be disclosed. Conversely, the coefficient will be negative if the proprietary cost hypothesis dominates and strong performers are hidden.

PROFITADJ is an industry abnormal profit adjustment measure similar to a construct used by Harris (1998). It captures the speed with which those industry participants with above-average profits have their positive abnormal profitability revert to the industry mean. We estimate *PROFITADJ* for each industry using the following industry-level panel regression over the period 1984 to 1997:[10]

[9] Unless otherwise noted the term industry refers to a group of firms, establishments, segments or pseudo-segments that have the same four-digit SIC code.

[10] In order to increase the precision of our estimates we estimate equation (2) on the entire panel of data. However, this suggests that the manager making the disclosure decision in 1987 or 1992 can predict the profit persistence in future years. Our conclusions are unchanged if we relax this assumption and estimate equation (2) using only historical data (i.e., for 1987 using 1984-1987, for 1992 using 1984-1992, and for 1997 using the entire panel).

$$X_{ijt} = \beta_{0j} + \beta_{1j} \times \left(DN_{ijt-1} \times X_{ijt-1}\right) + \beta_{2j} \times \left(DP_{ijt-1} \times X_{ijt-1}\right) + \varepsilon_{ijt} \qquad (2)$$

In equation (2), X_{ijt} is the year t difference between plant i's gross margin percentage per the LRD and the average gross margin percentage of all establishments in plant i's four-digit industry (i.e., industry j). DN_{ijt-1} is an indicator variable that equals one (zero) if X_{ijt-1} is (is not) less than or equal to zero. DP_{ijt-1} is an indicator variable that equals one (zero) if X_{ijt-1} is (is not) greater than zero.

The estimated coefficient on $DP_{ijt-1} \times X_{ijt-1}$ (i.e., β_{2j}) is used to measure the speed of adjustment for positive abnormal gross margin in industry j; hence, when β_{2j} is relatively high, abnormal profits adjust relatively slowly. For each pseudo-segment in our sample we set *PROFITADJ* equal to the estimate of β_{2j} for the industry that contains that pseudo-segment. While prior literature documents a negative relation between *MATCH* and *PROFITADJ*, the interpretation of this result is unclear. Harris (1998) interprets the negative coefficient as a manifestation of proprietary costs: firms aggregate pseudo-segments from industries where firms with above-average profits maintain their profitability advantage for long durations. However, a negative coefficient is also consistent with agency costs. That is, rather than the hidden pseudo-segments predominantly representing businesses with better profitability than their industry peers, firms may hide operations with below-average performance from shareholders because these operations are in an industry where other firms are able to consistently achieve above-average performance.

PRIVATE is a proprietary cost proxy. In the U.S., privately held firms have no obligation to publicly reveal financial results, whereas firms with publicly-traded debt or equity are required to file financial statements with the SEC. These financial statements are publicly available, providing potentially valuable information to competitors, suppliers, and others. Thus, we predict that the publicly-traded firms in our sample that compete in industries with high concentrations of privately-held firms will tend to mimic the disclosure policies of their private competitors.[11] We classify a Census establishment as

[11] A related interpretation of *PRIVATE* is that it is inversely related to barriers to entry in the industry, which may also be related to proprietary incentives to withhold information. Specifically, if industries with more privately-held firms require less capital to enter (hence the ability of private firms to stay private and avoid the capital markets), the

privately-held if it cannot be linked to a Compustat firm.[12] We then calculate *PRIVATE* as the ratio of total private-establishment *TVS* to total establishment *TVS* within the pseudo-segment's industry for that census year. We predict a negative association between *PRIVATE* and the likelihood of pseudo-segment disclosure.

TRANSIN is an indicator variable that equals one (zero) if the pseudo-segment receives (does not receive) inefficient transfers from the remainder of the firm. To calculate TRANSIN we begin by estimating transfers made or received by the pseudo segment. To do this we follow Rajan et al. (2000) and set transfers made (if negative) or received (if positive) equal to the difference between the actual investment made by the pseudo-segment and the investment the pseudo-segment would have made had it not been part of a multi-pseudo-segment firm (i.e., as if it were on its own). To estimate the investment the pseudo-segment would have made if it were on its own, we use the *investment ratio* of the remaining LRD plants (other than those in the pseudo-segment being assessed) in the pseudo-segment's four-digit SIC code. We define the investment ratio as the average of the ratio of net new plants (i.e., plants added less plants removed) to end-of-period plants. Hence, we arrive at the industry-adjusted investment ratio for pseudo-segment j ($IAIR_j$), which is shown in (3) below.

$$IAIR_j = \frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} \tag{3}$$

In equation (3) *NP* refers to net new plants added during the five year period, *EP* refers to the number of plants at the end of the five year period, *j* indexes pseudo-segments, and *i* indexes industries.

---

incumbent publicly-traded firms have an incentive to withhold disclosure from potential new entrants. We separately capture low barriers to entry in one of our other explanatory variables, *LOENT_HISUB*, and we find that *PRIVATE* has only a modest positive correlation with *LOENT_HISUB*. Thus, we view *PRIVATE* as primarily capturing the extent to which required disclosure by competitors is low, rather than the extent to which entry barriers are low.

[12] This variable contains three potential sources of measurement error. First, the matching of Compustat observations to observations in the Census' LRD is imperfect. In particular, these matches are based on the Census' Compustat-LRD Bridge File, which contains non-permanent Compustat identifiers taken from an edition of Compustat that is older than the one we are using. Second, some of the plants that we classify as private may be owned by non-U.S. *publicly*-traded firms that make information available as part of their home-country security laws and accounting principles. Finally, we ignore sales attributable to plants located outside the U.S. when calculating the ratio.

Our sample of publicly-traded firms likely contains firms with more funds available than the typical firm included in the LRD. In particular, many of the LRD firms that are not in our sample are small, privately-held companies; thus, they likely have relatively high costs of capital. Hence, for our sample firms, $IAIR_j$ likely overstates the amount transferred between pseudo segments. To correct for this, we follow Rajan et al. and subtract the industry-adjusted investment ratio averaged across the pseudo-segments of the firm from the pseudo-segment's industry-adjusted investment ratio. We refer to this industry- and firm-adjusted investment ratio as the *industry-firm adjusted investment ratio ($IFAIR_j$)*. This variable measures the transfers made (if negative) or received (if positive) by pseudo-segment j and is computed as follows:

$$IFAIR_j = \frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} - \sum_{j=1}^{n} w_j \left(\frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} \right) \tag{4}$$

In equation (4) n is the number of pseudo-segments in the firm and w_j is pseudo-segment j's share of total firm sales.

Finally, we set *TRANSIN* equal to zero for pseudo-segment j if: (1) $IFAIR_j$ is negative (i.e., pseudo-segment j is *making* transfers) or (2) $IFAIR_j$ is positive but the transfer received is deemed efficient. On the other hand, if $IFAIR_j$ is positive *and* the transfer received is deemed inefficient, *TRANSIN* is set equal to one for pseudo-segment j. We determine whether to deem a transfer as efficient or inefficient by comparing the industry Tobin's q of pseudo-segment j to the average q for all other pseudo-segments of the firm. If the transfer recipient's q is below (above) the firm average, then the transfer is deemed inefficient (efficient).[13] We predict a negative association between *TRANSIN* and the likelihood of pseudo-segment disclosure. In particular, based on the agency cost hypothesis, we expect that firms that

[13] We compute q ratios for 1985, 1990 and 1995 for each single-segment Compustat firm in every four-digit SIC code that has at least one Compustat segment in our sample in that year. Our q ratios are computed using the Lindenberg and Ross (1981) methodology and the specific assumptions of Hall et al. (1988). We then assign to each pseudo-segment for each five-year investment window the industry median q ratio of the Compustat single-segment firms that operate in the same industry as the pseudo-segment. We use the finest industry match that provides at least 5 single-segment Compustat firms with which to calculate the industry median q.

overinvest in low-growth-opportunity pseudo-segments are less likely to separately disclose these pseudo-segments in their external reports.

LOENT_HISUB is an indicator variable that measures proprietary costs of disclosure related to the extent of competition in the pseudo-segment's industry. Analytical predictions on the relation between discretionary disclosure and industry competition are complex because they depend on the nature of the competition, including whether the competition is from potential entrants or incumbents, whether firms compete on setting quantity or price, and whether the discretionary disclosure decision is an ex-ante commitment before the news is known or an ex-post opportunity after the news is revealed. Thus, one-dimensional measures of industry competition such as concentration ratios do not have clearly predictable relations to discretionary disclosure. We therefore construct *LOENT_HISUB* as a multidimensional measure of industry competition.

The *LOENT* portion of the variable name refers to low barriers to entry. Firms operating in product markets where entry costs are low face more competition from potential entrants. The *HISUB* portion of the variable name refers to high product substitutability. Firms operating in markets where products are highly substitutable face more competition from incumbents. We use the methods described in Tang (2009) to measure both of these phenomena. In particular, we use the ratio of total capital expenditures for a particular industry deflated by total revenues for that industry as our measure of barriers to entry. We measure product substitutability as the ratio of total revenue for a particular industry divided by the sum of raw materials and payroll costs for that industry. Low values of the entry barrier ratio combined with high values of the product substitutability ratio indicate the pseudo-segment faces high competition from both potential entrants and incumbents. Hence, we set *LOENT_HISUB* equal to zero *unless* the former ratio is below the median value of our sample industries *and* the latter ratio is above the median value of our sample industries, in which case we set *LOENT_HISUB* equal to one. If competition discourages discretionary disclosure, *LOENT_HISUB* will have a negative association with the likelihood of pseudo-segment disclosure.

LABOR measures the ratio of total labor costs to total sales revenues for a particular industry and it is thus measured, using LRD variables, as the ratio of total industry *SW* to total industry *TVS*. *SW* denotes salaries and wages and *TVS* denotes total value of shipments. While extensive analytical and empirical study has been devoted to the impact of product market competition on the proprietary costs of disclosure, little attention has been given to proprietary costs of disclosing information to customers or suppliers. *LABOR* captures proprietary costs of disclosing information to suppliers of labor under the maintained assumptions that these proprietary costs are greater when labor captures a large fraction of the firm's value added (Scott 1994; Liberty and Zimmerman 1986). If management wishes to obfuscate the true performance of the firm in order to maintain an information advantage over labor, there will be a negative association between *LABOR* and *MATCH*.

3.3.2. Control Variables

Firm size is an important variable in many settings and may be associated with discretionary disclosure policy for a number of reasons. Prior research indicates that greater size is generally associated with a higher level of disclosure (e.g., Lang and Lundholm 1993). We measure *FSIZE* as the natural logarithm of total firm assets (in $ millions) from Compustat.

We include the industry four-firm concentration ratio, *FOURFIRM*, to control for industry concentration. This ratio equals the fraction of aggregate industry *TVS* that is attributable to the largest four establishments (in terms of *TVS*) in that industry. Higher values of *FOURFIRM* imply higher levels of industry concentration. We match the industry measure to each pseudo-segment by four-digit SIC code to obtain the pseudo-segment's industry *FOURFIRM* measure. Harris (1998) uses this ratio calculated from Compustat data as one of her measures of competition and finds that it is negatively associated with the likelihood of disclosure. Ali et al. (2009), however, find no association between the ratio and the decision to provide segment disclosures when they use Census-based concentration measures. Thus, we offer no prediction on the association between *FOURFIRM* and *MATCH*.

The next three control variables, *RELSIZE*, *SEGDIVERSITY*, and *NUMEST* capture the firm's ability to aggregate the pseudo-segment within financial statement segments. *RELSIZE* is the ratio of pseudo-segment *TVS* to firm *TVS*. We expect a higher likelihood of pseudo-segment disclosure if the pseudo-segment is a relatively bigger component of firm operations; hence, we predict a positive association between *RELSIZE* and *MATCH*. *SEGDIVERSITY* is a measure of diversity. Under Statement 14, firms with operations in similar industries were afforded greater discretion to aggregate segment information than those with operations in diverse industries. We measure segment diversity as the ratio of the number of unique two-digit SIC codes across pseudo-segments to the total number of pseudo-segments (i.e., unique four-digit SIC codes). We predict a positive association between *SEGDIVERSITY* and *MATCH*. Finally, as an additional measure of firm complexity we include a count of the number of pseudo-segments that the firm operates in, *NUMEST*. We expect that as the count of pseudo-segments increases, aggregation of information is more likely. We expect a negative association between *NUMEST* and *MATCH*.

Recall that we eliminate all observations where firm-level *TVS* per the Census is either less than 75 percent of firm-level Compustat sales or exceeds 125 percent of firm-level Compustat sales. While this ameliorates measurement error, it does not eliminate it; hence, we include the variable *CEN_CMPSTAT*, which equals the ratio of firm-level *TVS* per the Census to firm-level sales per Compustat.

Firms report a numeric four-digit SIC code on their Census survey for the establishment's industry. However, when reporting externally in their annual reports they use a text-based description of the segment that *S&P* converts to a four-digit SIC code for Compustat. We use two variables that capture economic forces related to industry membership and that also control for possible measurement error in the *S&P* coding. First, we calculate *INDMATCHRATE*, which equals the proportion of pseudo-segments in an industry where *MATCH* equals one, excluding the sample pseudo-segment being analyzed. *INDMATCHRATE* captures the measurement error that results if *S&P* systematically misclassifies certain industries. Thus, we predict a positive association between *INDMATCHRATE* and *MATCH*.[14]

[14] This variable also captures a strategic aspect of segment disclosure. When an industry has a lower proportion of its establishments being separately disclosed as external segments, the proprietary costs of disclosing an establishment from that industry as a separate segment are likely higher. The proprietary cost hypothesis therefore also leads us to predict a positive relation between *INDMATCHRATE* and *MATCH*.

We also create an indicator variable called *NEC* for any pseudo-segment with a four-digit SIC code that is described as "Not Elsewhere Classified" in the SIC manual. For example, SIC 2389 is listed as "Apparel and Accessories, Not Elsewhere Classified." These are potentially idiosyncratic operations where *S&P* might err when converting a text description to a numerical one. Even if *S&P* does not err, it may be easier for firms to argue that an idiosyncratic operation is not part of a separate industry, but instead should be aggregated into a broader reported segment. We predict a negative association between *NEC* and *MATCH*.

4. EMPIRICAL RESULTS: MANAGERS' SEGMENT REPORTING DECISIONS

4.1. Descriptive Statistics and Sample Distribution

In Panel A of Table 1 we provide details about the observations lost at each of the steps in our sample selection and data validation processes. We begin with 9,975 Compustat observations from the years 1987, 1992 and 1997. We lose 4,195 observations that either cannot be linked to an observation in the Census's LRD (3,072 observations) or that have incomplete Census data (1,123 observations).[15,16]

An additional 2,667 observations that are on both Compustat and the LRD in 1987, 1992 or 1997 are eliminated because the observation does not meet our requirement of having at least two pseudo-segments. Finally, we remove 323 observations where firm-level *TVS* per the LRD exceeds 125 percent of firm-level sales per Compustat and 1,165 observations where the firm-level *TVS* per the LRD is less

[15] Failures to match Compustat to LRD data occur for at least two reasons. First, we are using Compustat data that were matched to the LRD by Census Bureau staff on the basis of name and address. In many cases, names in the Census data represent divisions and not ultimate parents and thus the firm may not be matched. Second, the original match was completed by Census in 2001 and CUSIP was used as the identifying variable in Compustat. If after 2001, the firm changed its CUSIP (or *S&P* reassigned the CUSIP), the firm will not match an observation in the edition of Compustat that we use.

[16] We compare the Compustat data median sales (assets) data for matched and unmatched firms and we find (in untabulated results) that the matched firms are more than three (two) times larger than unmatched firms. Matched firms have median sales (assets) of 102.5 (82.2) million dollars, while unmatched firms have median sales (assets) of 32.8 (39.2) million dollars. In addition to being smaller, we also find that the unmatched firms are less profitable (median return on sales of 2.7 percent versus 4.0 percent for the matched observations). The fact that unmatched firms are smaller corresponds to a finding reported by Maksimovic, Phillips and Prabhala (2008). They limit their sample to the Annual Survey of Manufacturers (*ASM*), however, which has limited coverage of firms with a small number of employees, whereas we use the comprehensive Census of Manufacturers (*CM*).

than 75 percent firm-level sales per Compustat. These steps result in a final sample of 1,625 observations (representing 1,008 unique firms).

Panel B of Table 1 contains descriptive statistics about the degree of disaggregation. The average number of Census pseudo-segments per firm, at 5.1, is more than three times the corresponding average of 1.6 Compustat segments per firm and the standard deviation of pseudo-segments per firm is more than five times the standard deviation of segments per firm. Thus, even after grouping plants together by four-digit SIC code, the Census data exhibit a higher level of disaggregation than the Compustat segment data and a higher degree of variability in the extent of disaggregation.

In Panel C of Table 1 we present the sample's distribution by two-digit SIC code and compare our sample's industry distribution to that of the two-digit manufacturing SIC codes on Compustat.[17] The first two columns of numbers in Panel C show that pseudo-segments are fairly evenly distributed across industries, with only SIC codes 34 (Fabricated Metals & Transportation Equipment), 35 (Industrial/Commercial Machinery & Computers) and 36 (Electrical Equipment) having more than 10 percent of our sample's pseudo-segments. In contrast, *TVS* is somewhat concentrated in combined SIC codes 20 and 21 (Food and Tobacco Products) and in SIC code 37 (Transportation Equipment).

The middle two columns of numbers in panel C contain the industry distribution for Compustat segments reported by firms for which at least 75 percent of the firm's sales come from manufacturing industries, whereas the final two columns of the panel present the industry distribution for all Compustat segments that have primary or secondary SIC codes between 2000 and 3999. The two sets of Compustat figures are very similar and reveal that, although the industry distribution of our sample is broadly similar to that of the Compustat population, our sample is underrepresented in the Chemicals & Petroleum industry and overrepresented in the Fabricated Metals & Transportation Equipment industry.

[17] Census disclosure requirements caused us to combine two industries in all situations where one of the industries exhibited fairly small representation.

4.2. Univariate Statistics

In Table 2 we examine how descriptive statistics for our key explanatory variables differ between the observations for which the dependent variable, *MATCH*, equals one and those for which *MATCH* equals zero. The unit of analysis in this table, and generally for the remaining tables, is the pseudo-segment. We divide Table 2 into three panels that capture three different samples.

The full sample (Panel A) includes all 1,625 firm-year observations, and thus all 8,287 of our sample pseudo-segments. The single-segment sample (Panel B) contains the 956 firm-year observations (with 3,300 pseudo segments) for which the firm reports either one line-of-business segment or has no line-of-business segment footnote in its 10-K filing. Finally, the multisegment sample (Panel C) contains the 669 firm-year observations (with 4,987 pseudo-segments) in which the firm reports two or more line-of-business segments. In each of the three samples, several major differences in the values of our explanatory variables emerge between the observations for which *MATCH* equals one and those for which *MATCH* equals zero. All differences discussed below (both means and medians) are statistically significant at the one percent level unless otherwise stated.

For the full sample results presented in Panel A, approximately 37 percent of the pseudo-segments are disclosed as Compustat segments (i.e., *MATCH* equals one for 3,032 of 8,287 observations). The median value of *I_PROFIT* is 0.08 for undisclosed pseudo-segments and 0.11 for disclosed pseudo-segments.[18] This represents a 38 percent difference and is consistent with the agency cost motive as it suggests that pseudo-segments are more likely to be aggregated when abnormal profitability is lower.

PROFITADJ is the *industry* rate of positive abnormal profit adjustment, where profit is based on gross margin percentage.[19] The higher mean and median values of *PROFITADJ* for the non-disclosing

[18] The fact that *I_PROFIT* is positive for both disclosed and undisclosed pseudo-segments indicates that gross profits of the sample firms' pseudo-segments exceed those of their industry peers, consistent with the publicly-traded firms that enter our sample being more profitable than the (generally) privately-held firms that are excluded from our sample.

[19] The mean coefficient estimates of 0.32 for disclosed pseudo-segments and 0.35 for nondisclosed pseudo-segments are both considerably lower than Harris's (1998) cross-sectional industry average (0.70 per her Table 1). Differences arise for at least three reasons: (1) we sample from different time periods than Harris (1987, 1992 and 1997 for our sample versus 1987-1991 for Harris); (2) we use different databases (Census LRD for our sample,

sample indicate a slower speed of positive abnormal profit adjustment for the industries of non-disclosing pseudo-segments, consistent with the findings in Harris (1998), Botosan and Stanford (2005), and Berger and Hann (2007).

The seven percent higher mean and median values of *PRIVATE* for the non-disclosing sample indicate that pseudo-segments that are not disclosed are from industries with a relatively high concentration of private firms. The mean values of *TRANSIN* indicate that 32 percent of undisclosed pseudo-segments receive inefficient transfers whereas only 26 percent of disclosed pseudo-segments are recipients of such subsidies. Thus, consistent with the agency cost hypothesis, the univariate statistics indicate that pseudo-segments that receive subsidies are less likely to be disclosed.

LOENT_HISUB has a mean value of approximately 28 percent for both disclosed and undisclosed pseudo-segments. However, the mean and median values of *LABOR* are about one percent higher at undisclosed pseudo-segments, consistent with disclosure being less likely when salaries and wages represent a larger fraction of sales in the pseudo-segment's industry.

For brevity, we do not discuss the statistics related to our control variables, with the exception of *RELSIZE*, which is the ratio of pseudo-segment *TVS* to firm-level *TVS*. Untabulated results reveal that the median value of *RELSIZE* is 0.08 across the combined sample, indicating that the median pseudo-segment has sales that represent 8 percent of firm sales. Thus, the majority of the pseudo-segments represent a small enough portion of the firm to easily be aggregated within a financial statement segment under SFAS 14 rules that require an industry segment to be separately reported if the segment's revenue, earnings, or assets are at least 10 percent of the combined total for that item across all of the firm's industry segments.[20]

---

which includes private firms in each industry versus Compustat for Harris, which excludes most private firms); and, (3) we use different profit measures (gross margin percentage for our study versus ROA in Harris).

[20] More precisely, paragraph 15 of SFAS 14 requires an industry segment to be classified as a reportable segment if it satisfies one or more of the following tests when the tests are applied separately for each fiscal year for which financial statements are presented: (a) Its revenue is 10 percent or more of the combined revenue of all of the enterprise's industry segments. (b) The absolute amount of its operating profit or operating loss is 10 percent or more of the greater, in absolute amount, of: (i) The combined operating profit of all industry segments that did not incur an operating loss, or (ii) The combined operating loss of all industry segments that did incur an operating loss. (c) Its identifiable assets are 10 percent or more of the combined identifiable assets of all industry segments. SFAS 14

Panel A of Table 2 shows that the median value of *RELSIZE* is 0.26 for disclosed pseudo-segments versus 0.04 for those not separately disclosed as financial statement segments. This indicates that disclosed pseudo-segments generally contribute more than 10 percent of firm sales, a trigger point for separate segment disclosure under SFAS 14, whereas undisclosed pseudo-segments generally do not. Thus, it is critical to control for *RELSIZE* in our multivariate tests. Nevertheless, the large standard deviation for *RELSIZE* of 0.17 among undisclosed pseudo-segments indicates that a considerable portion of pseudo-segments with *RELSIZE* above 0.10 do not get disclosed as separate segments. The standard deviation of *RELSIZE* is 0.30 for disclosed pseudo-segments, which suggests that many with *RELSIZE* below 0.10 *do* get disclosed separately. Thus, while important, *RELSIZE* is not deterministic with regard to the pseudo-segment disclosure decision.

The results in Panels B and C of Table 2 reveal that the differences between the disclosing and non-disclosing pseudo-segments documented in panel A also hold for the single-segment and multisegment sample with one exception: for the single-segment sample *PROFITADJ* is no longer statistically significantly larger for the non-disclosing pseudo-segments.

Table 3 includes a correlation matrix for the multisegment sample with Pearson correlations above the diagonal and Spearman below.[21] Because our discussion of Table 2 covers the main associations between our dependent variable, *MATCH*, and the hypothesized explanatory variables, we do not repeat those associations here in the context of the Table 3 correlations. We do note from Table 3, however, that most of the explanatory variables have significant correlations not only with *MATCH*, but also with each other. We also note that *RELSIZE* has a large positive correlation with *MATCH* as well as statistically significant (although much smaller) correlations with our key inferential variables. As we discuss in detail later, the relation between *RELSIZE* and *MATCH* also has an important nonlinearity due to the inclusion of a cutoff point for measures similar to *RELSIZE* in the accounting rules for segment reporting.

---

goes on in paragraph 16 to state that interperiod comparability could require an industry segment to (not) be reportable even if it falls below (above) the 10 percent cut-offs in the currently reported periods.

[21] For brevity, we present correlations only for the multisegment sample because the majority of our multivariate analyses focus on this group.

4.3. Base Model Logistic Regression Analysis

4.3.1. Full Sample

Table 4, Panel A presents the results of estimating our base model logistic regression on the full sample. The first column of numbers presents the coefficient estimates and the pseudo R-squared of the logistic regression. The pseudo R-squared exceeds 20% in Panel A of Table 4 and in all of our subsequent logisitic regressions, indicating reasonable goodness of fit for our regression models. The second column presents the two-tailed p-values, and the third a measure of the economic magnitude of the variables' effects (the marginal probability effect of the variable with all other explanatory variables evaluated at their mean, multiplied by one standard deviation of the variable, or simply by one if it is an indicator variable). Because we include multiple observations for the same firm in our regressions, p-values are calculated using robust standard errors that correct for firm clustering.

Before we turn to the estimates for the main variables, the results for several of the control variables are worth noting. First, the coefficient estimate on *RELSIZE* is positive and extremely significant both statistically and economically, consistent with pseudo-segments that represent a larger portion of their firm's sales being more likely to be disclosed. This is not surprising, as smaller pseudo-segments are probably easier to "hide" as they are less visible to investors and competitors, and as a result of the size threshold for segment reporting in SFAS 14.

We draw some comfort from the fact that the estimate on *CEN_CMPSTAT* is not significantly different from zero. This indicates that including firm-years where Census *TVS* captures as little as 75 percent or as much as 125 percent of the total firm sales per Compustat does not distort our inferences about which Census pseudo-segments are disclosed as Compustat segments. One variable that mitigates S&P measurement error, *INDMATCHRATE*, is highly significant, while the other, *NEC*, is not (although it does have the predicted sign). The positive coefficient on *INDMATCHRATE* suggests that a pseudo-segment is more likely to be disclosed when other firms' pseudo-segments in the same industry tend to be disclosed. *NUMEST* captures the number of pseudo-segments at the firm level, and it is negatively associated with the likelihood of disclosing an individual pseudo-segment. This is intuitive, as more pseudo-segments increase the firm's ability to aggregate.

The control variable *FOURFIRM* is not related to *MATCH*. Recall that Harris (1998) documents a negative association between this industry concentration measure and the probability of disclosure. One reason for the difference between our result and Harris's is that she uses a Compustat-based measure of industry concentration. Our measure, on the other hand, is based on Census data that include private firms in each industry. Hence, we avoid the issues discussed in Ali et al. (2009), who demonstrate that industry concentration measures constructed with Compustat data have important limitations attributable to Compustat's exclusion of privately-held firms.

In summary, before turning to our hypothesis tests, we conclude that our dependent variable, *MATCH*, which captures the internal pseudo-segment disclosure aggregation decision, is reasonably associated with firm fundamentals that are likely to be "natural" (i.e., non-strategic) determinants of the disclosure decision.

The six treatment variables of interest are *I_PROFIT*, *PROFITADJ*, *PRIVATE*, *TRANSIN*, *LOENT_HISUB*, and *LABOR*. The full sample results that follow use the combination of the single-segment and the multisegment samples, which we argue need to be evaluated separately. Thus, while these results are presented for completeness, they are not a major source of our final inferences.

We find that the coefficient on *I_PROFIT* is positive and significant. This indicates that, on average, pseudo-segments with high abnormal profitability are more likely to be disclosed, which is consistent with the agency cost hypothesis that managers are more likely to withhold information about pseudo-segments with low abnormal profitability.

The coefficient on *PROFITADJ* is negative, but it is insignificant in the full sample. This result is inconsistent with Harris's (1998) finding of a negative and significant association. The significant, negative coefficient on the variable *PRIVATE* is consistent with the likelihood of pseudo-segment disclosure being reduced when the pseudo-segment competes in an industry that has a large concentration of privately-held firms. The significantly negative coefficient estimate on *TRANSIN* is consistent with the agency cost hypothesis of pseudo-segments being less likely to be disclosed when they are receiving inefficient transfers. Finally, the results for both *LOENT_HISUB* and *LABOR* are insignificant both statistically and economically.

4.3.2. Single-segment Firms

As discussed in the introduction, an advantage of the census data is that they give us the opportunity to analyze aggregation decisions made by managers of single-segment firms, which is a distinctive feature of our study. For example, Harris (1998) and Botosan and Harris (2000) evaluate only multisegment firms. Berger and Hann (2007) and Botosan and Stanford (2005) do examine firms that reported one segment under SFAS 14 that was restated to multiple segments under SFAS 131, but these restating firms represent a minority of the total population of single-segment firms.

Single-segment firms are of interest because it is likely that they differ from multisegment firms in two related ways. First, from a strategic disclosure standpoint, line-of-business aggregation has fundamentally different consequences for single-segment firms than for multisegment firms. If the firm chooses to report a single external segment, outsiders know for certain that *all* industries the firm operates in are aggregated into that segment.[22] On the other hand, if the firm chooses to report multiple external segments, an additional layer of uncertainty is introduced. In this situation, there are now multiple reported divisions that might contain several possible combinations of aggregated (i.e., non-disclosed) business activities. Because this additional layer of strategy is not applicable to single-segment firms, we predict that strategic factors play a lesser role in determining the aggregation decisions of single-segment firms vis-à-vis multisegment firms. Second, we expect that agency costs play a less pertinent role in determining the aggregation decisions of single-segment firms. We base this prediction on past research, which shows that, on average, multisegment firms are traded at a discount relative to single-segment firms (Lang and Stulz 1994; Berger and Ofek 1995).[23]

---

[22] That said they will not know about the profitability of individually aggregated activities, so some uncertainty remains.

[23] These expectations are reasonable only if SFAS 14 imposed some limits on the ability of firms with many pseudo-segments to report only one financial statement segment. There is evidence that suggests these limits existed and,

We present the results of our multivariate analyses of the single-segment sample in Panel B of Table 4.[24] Estimating the equation (1) base model on the single-segment sample produces substantially different results relative to those reported in Table 4, panel A. In contrast to the full-sample results, only one of the treatment variables, *PRIVATE*, is associated with the likelihood of disclosure. Hence, there is no evidence that agency costs motivate the disclosure decisions of managers of single-segment firms. Rather, it appears that proprietary costs arising from private competition are the key strategic determinant of pseudo-segment aggregation for these firms.

4.3.3. Multisegment Firms

Table 4, Panel C presents the results of estimating our base model on the multisegment firms. Before discussing specific results it is worth noting that the overall results are considerably different from those shown in either Panel A (i.e., the full-sample results) or Panel B (i.e., the single-segment sample results). In particular, for the multisegment sample, four of the treatment variables we consider have a significant association with *MATCH*; moreover, *PRIVATE* is not one of these variables. However, as discussed above, *PRIVATE* is the only treatment variable that is related to the aggregation choices of single-segment firms. These results suggest that compared to single-segment firms: (1) strategic motives play a larger role in the determination of multisegment firms' aggregation choices and (2) agency cost motives for nondisclosure are more pertinent for multisegment firms. These results also suggest that the proprietary costs that are relevant in the multisegment firm context have a different form than those that are relevant to managers of single-segment firms.

Turning to the individual results, we find that *I_PROFIT* has a positive and significant coefficient estimate. This implies that, on average, pseudo-segments with high abnormal profitability are more likely to be disclosed. The result is consistent with the agency cost hypothesis that pseudo-segments with low

---

thus, SFAS 14 single-segment firms truly were less diversified than their multisegment peers. In particular, as show in Panel B of Table 1 of Berger and Hann (2003), 77 percent of SFAS 14 single-segment firms continue to report only one external segment under SFAS 131, which arguably reduced managers' ability to aggregate.

[24] In this specification, we remove the control variables *SEGDIVERSITY* and *NUMEST* because we expect that once a firm decides to report only one segment, the diversity and number of industries have no effect on the identity of the business activity that is reported as the firm's primary operations. None of our conclusions change when we include these variables in the model, and neither is significant when included.

abnormal profitability are more likely to be aggregated by multisegment firms. The economic magnitude of *I_PROFIT* is also significant: a one standard deviation increase in *I_PROFIT* increases the probability of pseudo-segment disclosure by 2.8 percent.

The *PROFITADJ* variable is also negative and significant in the multisegment sample. This suggests that pseudo-segments are more likely to be aggregated when they operate in industries with persistent levels of abnormal profits, consistent with Harris (1998). The economic magnitude column shows that, for *PROFITADJ*, a one standard deviation increase from the mean value of 0.33 is associated with a 7.3 percent decrease in the probability of disclosure. Harris interprets this finding as consistent with the proprietary cost motive for aggregation. Such an interpretation implies that the slow convergence of the top performers toward the industry mean is driven by the stronger firms continuing to protect their proprietary advantages by using less disaggregated disclosure. An alternative possibility, forwarded by Berger and Hann (2007), notes this finding is also consistent with the agency cost motive for aggregation. Berger and Hann point out that in an industry with a high level of persistently positive abnormal profits, the slow convergence of the top performers toward the industry mean can be driven by the weaker firms continuing to experience unresolved agency problems and failing to fully disclose their underperforming operations (as opposed to the top firms having a competitive advantage via withholding proprietary information). We explore these alternative explanations in subsequent analyses.

Regarding *TRANSIN*, the significantly negative coefficient estimate is consistent with the agency cost hypothesis that pseudo-segments are less likely to be disclosed when they are receiving inefficient transfers. Here, again, the economic magnitude is meaningful: when the pseudo-segment is classified as receiving inefficient transfers, the probability of non-disclosure increases by 4.2 percent.

The coefficient on *LOENT_HISUB* is insignificant, which is inconsistent with product-market competition as reflected in low barriers to entry and high product substitutability affecting the segment disaggregation decision. Finally, the importance of labor costs in the industry is associated with the likelihood of pseudo-segments in that industry being separately disclosed. The significantly negative coefficient estimate on *LABOR* is highly important economically, with a one standard deviation increase in labor costs being associated with a 32 percent decrease in the probability of separately reporting the pseudo-segment.[25]

5. EXTENSIONS OF THE BASE MODEL AND SENSITIVITY ANALYSES

In this section we extend the base model shown in equation (1). Given the sharp differences in results from estimating equation (1) on the multisegment versus single segment samples, it is clear that mixing these samples together clouds the inferences that can be drawn from estimation of equation (1). Accordingly, we focus the remainder of the paper on the sample of multisegment firms.

5.1. Nonlinear Size Effects

A potential concern with the base model is whether using only a linearly additive control for the relative size of the pseudo-segment is sufficient. The accounting rules for segment reporting suggest that the impact of relative pseudo-segment size on the aggregation decision may be nonlinear; specifically, the impact may depend on whether the size of the pseudo-segment exceeds the 10 percent cutoff mentioned in paragraph 15 of SFAS 14 (see footnote 21 above). Therefore, we modify the base model by adding seven new variables. In particular, we add an additional fixed effect, *SML_RELSIZE*, which is an indicator variable set equal to one when *RELSIZE* is less than 0.10, and zero otherwise. The remaining six new variables are interactions between *SML_RELSIZE* and our six treatment variables.

The results from estimating this extended model are shown in Table 5.[26] Three main points emerge. First, the coefficient estimate on *SML_RELSIZE* is significantly negative. This indicates that even after

---

[25] We also estimate this model on a sample of multisegment firms that appear in all three years of the panel. This sample consists of 2,281 pseudo-segments, or approximately 46 percent of the original sample size. Our inferences are unchanged when using this specification, with the exceptions that while *I_PROFIT* and *LABOR* continue to have the same signs, the statistical significance of the estimated coefficients declines to p-values of 0.21 and 0.47.

[26] Research by Ai and Norton (2003) and Powers (2005) emphasizes that it can be potentially difficult to interpret interaction terms from non-linear models such as probit or logisitic regression models *if* one is interested in assessing marginal effects at a point other than the center of the distribution and *if* the researcher is interested in quantifying the marginal effect of the interaction term on the probability of an event (pseudo-segment disclosure in our setting). We do not believe that either of these circumstances apply to our Table 5 regression nor to the subsequent untabulated sensitivity tests that use the same interactions as in Table 5 (between *SML_RELSIZE* and our six treatment variables). Nevertheless, from an abundance of caution we further ensure that our inferences are correct by adopting the methodology developed by Norton, Wang and Ai (2004) to compute marginal effects of the interaction terms and assess these effects over the entire range of predicted probabilities from our logisitic regression

including the variable *RELSIZE*, there remains a significant "size effect." In particular, and as expected, the likelihood of disclosure declines significantly when the pseudo-segment's relative size drops below the 10 percent cutoff specified by the accounting rules. Second, the coefficient estimates from the main effects of our six treatment variables are the same in Table 5 as in Panel C of Table 4, with the exception that the estimate on *I_PROFIT* is no longer significant. Thus, pseudo-segments that have a relative size of at least 10 percent are affected by our treatment variables in the same manner inferred from the multisegment results of Table 4, with the exception that the agency cost hypothesis that more profitable pseudo-segments are more likely to be disclosed is no longer supported.

Third, the coefficient estimates on the interactions between the *SML_RELSIZE* indicator and our six treatment variables are insignificant, with the exception of the estimate on *LABOR*×*SML_RELSIZE*, which is significant at the 0.10 level. Thus, relative industry-level labor costs have little impact on the decision to separately disclose a pseudo-segment when the pseudo-segment is small (i.e., when it has *TVS* that is less than 10 percent of firm-level *TVS*).

Given these results, we conclude that while the disclosure decision is not linear with respect to pseudo-segment size (i.e., the main effect of *SML_RELSIZE* is significant), this nonlinearity does not have a first-order effect on our inferences (i.e., the interactions are almost all insignificant).

5.2. Effects of the Excess Value of Diversification

Our results suggest that the motive, means and opportunity to aggregate for agency cost reasons are greater for multisegment firms than for single-segment firms. Within the multisegment sample, however, there is likely considerable variation in the agency cost motive. To exploit this variation, we conduct additional tests by re-estimating our base regression on two subsamples of the multisegment sample. In particular, we separate pseudo-segments into two groups based on whether the pseudo-segment belongs to a firm with negative or nonnegative estimated excess-value from diversification. We use the methodology discussed in Berger and Ofek (1995) to estimate excess values from diversification. A

---

models. We find that our inferences are unambiguous across the entire range of predicted probabilities for all of the models that contain interactive terms.

negative excess value from diversification indicates that the value of the multisegment firm is lower than the value of a collection of stand-alone firms consisting of the multisegment firm's separate lines of business. To the extent this value loss is a manifestation of agency problems, negative excess value firms are more likely to have agency costs than firms with nonnegative excess values.

Table 6, Panel A (B) shows the results from estimating our base model on the sub-sample of multisegment firms that have negative (nonnegative) excess values. Of the 4,987 pseudo-segments available for the logistic regressions in Tables 4 and 5, the requisite data for calculating excess values are available for 4,517. Of these, 2,545 (56%) pseudo-segments are from firm-years with negative excess values and the remaining 1,972 (44%) pseudo-segments are from firm-years with nonnegative excess values. Overall, for all multisegment firms, untabulated results show that the mean excess value is -0.13 and the median is -0.08; this implies that the diversification discount identified in past literature on broad samples (Berger and Ofek 1995; Lang and Stulz 1994) exists in our sample of firms concentrated in domestic manufacturing.

With regard to the proprietary cost variables, neither *PRIVATE* nor *LOENT_HISUB* are associated with *MATCH* in either panel A or panel B. Hence, we continue to find no evidence that either of these factors affect multisegment firms' pseudo-segment aggregation decisions. The results for *PROFITADJ* do, however, offer support for the proprietary cost hypothesis because the significantly negative estimate on this variable that we observed in Tables 4 and 5 is attributable to the sub-sample of multisegment firms that have nonnegative excess values. Given that positive excess value firms presumably do not have an agency cost motive to suppress disclosure, this finding is consistent with *PROFITADJ* capturing a proprietary cost motive for pseudo-segment aggregation.

Turning to the agency cost variables, we see that the estimates on *I_PROFIT* and *TRANSIN* are significant in the predicted direction for the negative excess value subsample presented in Panel A. Thus, as we expected, firms with value-reducing diversification strategies, and arguably higher agency costs, are less likely to separately disclose pseudo-segments that are less profitable, as well as those that receive more inefficient transfers from other pseudo-segments. However, as shown in Panel B, the coefficient on

TRANSIN is also significantly negative for the nonnegative excess value subsample. Moreover, as shown in Panel A, the proprietary cost variable *LABOR* is significant in the negative excess value sub-sample but not the nonnegative excess value subsample, which is the opposite of our expectations. Thus, while the agency cost motive for aggregating pseudo-segments appears to apply strongly to the multisegment firms with unsuccessful diversification strategies, there are some inconsistencies with our expectations.[27]

5.3. Un-tabulated Sensitivity Analyses[28]

5.3.1. Validation and Sensitivity Testing of our Dependent Variable, *MATCH*

As discussed in section 3.2, we face a difficult research design issue in identifying the aggregation schemes firms *should* use if they neutrally apply the accounting principles mandating separate industry disclosure. The accounting standard in effect for the sample period (SFAS 14) explicitly rejects applying any single industry classification system to all firms yet, in order to conduct a meaningful and objective empirical study, as researchers we are forced to adopt a single system. In section 3.2 we defend, on an *a priori* basis, our use of four-digit SIC codes applied to pseudo-segments. In this section we provide empirical evidence that supports our arguments and we discuss the results of sensitivity analyses in which we use an alternative aggregation scheme that is based on three-digit SIC codes.

We begin by evaluating the construct validity of our disclosure measure, *MATCH*, using the data set collected by Berger and Hann (2003; 2007), who study firms that restate historical SFAS 14 segment data as part of their initial adoption of SFAS 131. There are 99 firms that overlap between our sample and theirs.[29] We calculate two ratios at the firm level, one from each data set. First, using Berger and Hann's

[27] We evaluate the robustness of the Table 6 results by estimating the extended version of equation (1) that incorporates *SML_RELSIZE* and its interactions with our six treatment variables rather than equation (1). The (un-tabulated) results show that the only finding for the negative excess value sample sensitive to using the expanded model is that the coefficient on *I_PROFIT* is no longer significant. For the nonnegative excess value sample, two results are noteworthy. First, the marginally significant, and negative, coefficient on *TRANSIN* found in Panel B is statistically insignificant in the expanded model. Thus, the somewhat surprising finding that receiving inefficient transfers reduces the likelihood of pseudo-segment disclosure at firms with value-enhancing diversification programs is not robust. Second, similar to the results shown in Table 5, the coefficient estimate on *LABOR*×*SML_RELSIZE* is positive and significant.

[28] These results are available, upon request, from the authors.

[29] We analyze 1,625 firm years across 1987, 1992 and 1997. Berger and Hann use 796 restating firms across 1997 and 1998. The overlap is reduced from 796 because we require our firms to be almost exclusively involved in

data, we identify the newly revealed segments in 1998 annual reports; these are segments that were not disclosed initially in 1997 under SFAS 14, but were subsequently revealed as restated data upon initial adoption of SFAS 131. For each firm, we divide these newly revealed segments by the total number of restated 1997 segments per the 1998 annual report. Second, using Census data, we identify the total number of unmatched pseudo-segments (i.e., *MATCH* equals zero) for each firm in 1997, and divide this by the total count of pseudo-segments for that firm-year.

Both of the ratios mentioned above are measures of non-disclosure, and they should be positively correlated. Empirically, we find that this is the case. Specifically, the Pearson correlation is 0.25 (p-value of 0.01) and the Spearman correlation is 0.22 (p-value of 0.03). Hence, our dependent variable *MATCH* exhibits external validity.

Next, we discuss results from using three-digit SIC codes to define industry in lieu of four-digit SIC codes. We do not believe this approach is appropriate for our primary analyses, as it adopts a disclosure benchmark that is too aggregated. That is, if users of financial statements demand detailed industry information, an industry classification scheme that aggregates more and more disparate types of operations is contradictory to their demands.[30]

We repeat our logistic regressions estimated for the firms that externally report multiple segments. We adjust the specification shown in equation (1) in three ways. First, we form pseudo-segments on the basis of three-digit SIC codes and we use a revised version of *MATCH* that equals one (zero) if a particular pseudo-segment matches (does not match) either the primary or secondary three-digit SIC code of any of the firm's externally reported segments. Second, we define industries on the basis of three-digit SIC codes when measuring all of our industry-adjusted (i.e., *I_PROFIT* and *TRANSIN*) and industry-level (e.g., *PROFITADJ* and *PRIVATE*) variables. Finally, we exclude the variable *NEC*, which is designed to

---

domestic manufacturing, whereas Berger and Hann have neither the industry nor geographic limitation, and they sample from 1998 which we do not.

[30] In fact, the development of the more detailed NAICS suggests that even the four-digit SIC code system may be too coarse for users of industry data.

control for measurement error related to miscellaneous industries and/or industries that are difficult to classify. These issues are irrelevant when moving up to the three-digit SIC code level.

The (untabulated) results from the specification described above show that the treatment variables *PROFITADJ* and *TRANSIN* have strong, negative and highly-significant (p-values < 0.01) associations with *MATCH* even when industry definitions are based on three-digit SIC codes. Hence, the main results documented in the tables are robust to an alternative industry classification scheme. However, there are two pertinent differences between the untabulated results and the results presented in Panel C of Table 4: (1) the treatment variables *I_PROFIT* and *LABOR* are no longer significant and (2) the control variable *FOURFIRM* has a positive, significant association with *MATCH*.

5.3.2 Effects of Information Environment Variables

Our tests thus far have not controlled for information environment variables shown in prior research to be associated with various aspects of discretionary disclosure. In this robustness test we add four explanatory variables to our regressions. These variables capture aspects of the information environment associated with disclosure quality. These four variables are listed below:

- *FOLLOW* equals the average number of I/B/E/S analysts providing an estimate of the firm's annual earnings over the 12 months of the firm's fiscal year. Lang and Lundholm (1996) show that analyst following is higher at firms when disclosure practices are rated as more informative by committees of financial analysts.
- *ROS* denotes return on sales and it equals the firm's net income plus preferred share dividends divided by sales revenue in the same year that *MATCH* is measured. Prior research finds that many measures of disclosure quality are positively correlated with contemporaneous measures of firm performance (e.g., Lang and Lundholm 1993).
- *NETDEBTISSUED* is a flow-of-funds based measure of the firm's net debt issuance. We follow Berger, Ofek and Yermack (1997) and calculate it as debt issued minus debt retired, all scaled by total assets [i.e., Compustat annual data items (#111 - #114 + #301) / (#6)].

- *EQUITYISSUED* is a flow-of-funds based measure of new equity issued by the firm scaled by total assets [i.e., Compustat annual data items (#108) / (#6)].[31]

We include *NETDEBTISSUED* and *EQUITYISSUED* to control for incentives to improve disclosure in advance of accessing capital markets (Lang and Lundholm 1993; Frankel, McNichols and Wilson 1995); hence, we measure both of these variables in the year after the measurement of *MATCH*.

The (untabulated) results from estimating an extended version of equation (1) that includes the four variables described above lead to inferences that are qualitatively identical to those drawn from each panel of Table 4 with respect to our six treatment variables. With regard to the information environment variables themselves, the coefficient estimates on the debt and equity issuance variables are insignificant, as is the estimate on the analyst following variable; however, the estimate on the return on sales variable is always positive and highly significant.

We also estimate an extended version of the Table 5 model that includes each of the four information environment variables described above. The (untabulated) results from this regression lead to identical inferences to those drawn from the results shown in Table 5, except that the coefficient estimate on *LABOR* x *SML_RELSIZE* is no longer significantly different from zero at the .10 probability level.

6. CONCLUSION

We examine discretionary disclosure with novel data, which we exploit to examine a comprehensive set of forces hypothesized to influence segment reporting decisions. Our data consist of *confidential* information about internal operating results for a large sample of U.S. manufacturing firms. Our proxies capture various aspects of agency and proprietary cost motives for nondisclosure. The combination of unique data and a broad set of disclosure determinants allows us to provide new insights regarding the motives underlying managers' aggregation decisions. This is the major contribution of our study.

---

[31] By using Compustat flow-of-funds data to measure debt and equity issuance we are able to capture events (such as bank debt issuance or redemption) involving management's discretionary external funding transactions that are missed by databases of public securities issuance. On the other hand, a disadvantage of this approach is that it will also capture some external funding transactions triggered by events other than management's discretion, including outside claim holders exercising stock options or warrants, or converting convertible debt into equity.

We provide evidence that suggests there are nontrivial differences between the disclosure behavior of single-segment firms and the disclosure behavior of multisegment firms. In particular, single-segment firms exhibit less strategic behavior and, to the extent these firms strategize, our evidences suggests that proprietary costs related to private competition are the key motive for nondisclosure. On the other hand, our evidence suggests that both proprietary and agency motives are important determinants of multisegment firms' segment disclosure decisions.

Regarding the effect of agency costs on the disclosure choices of the multisegment firms in our sample, we show that if a pseudo-segment is receiving inefficient transfers of capital from the rest of the firm, it is less likely to be disclosed separately. This result is robust to alternative specifications. We also show that pseudo-segments with low industry-adjusted profits are more likely to be aggregated into another external segment and, thus, hidden from outsiders. This result is, however, sensitive to model specification.

With respect to proprietary costs, we find that if a multisegment firm has operations in an industry where positive abnormal profits of some industry members (relative to the industry average) are likely to persist, it is less likely to disclose those operations separately. This result is robust and it is primarily manifest in the sub-sample of firms least likely to have an agency cost motive for nondisclosure (i.e., those where diversification appears to be value enhancing for shareholders). This suggests that it relates to a proprietary cost effect, which is consistent with conclusions drawn by Harris (1998). We also find that multisegment firms are more likely to aggregate operations that are from an industry with greater labor power. This result is sensitive to model specification, however.

Finally, in addition to shedding light on the proprietary and agency cost motives for nondisclosure, we also show that comprehensive measures of industry statistics (i.e., including both privately-held and publicly-traded firms) are useful for measuring both industry-adjusted and industry-level determinants of disclosure.

REFERENCES

Ai, C. and E. Norton. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters* 80, 123-129.

Ali, A., S. Klasa and E. Yeung. 2009. The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research. *Review of Financial Studies* (forthcoming).

Bamber, L. and Y. Cheon. 1998. Discretionary Management Earnings Forecast Disclosures: Antecedents and Outcomes Associated with Forecast Venue and Forecast Specificity Choices. *Journal of Accounting Research* 36*,* 167-190.

Bens, D. and S. Monahan. 2004. Disclosure Quality and the Excess Value of Diversification. *Journal of Accounting Research* 42*,* 691-730.

Berger, P. and R. Hann. 2003. The Impact of SFAS No. 131 on Information and Monitoring. *Journal of Accounting Research* 41*,* 163-223.

______________. 2007. Segment Profitability and the Proprietary and Agency Costs of Disclosure. *The Accounting Review* 82, 869-906.

Berger, P. and E. Ofek. 1995. Diversification's Effect on Firm Value. *Journal of Financial Economics* 37, 39-65.

______________. 1999. Causes and Effects of Corporate Refocusing Programs. *Review of Financial Studies* 12, 311-345.

Berger, P., E. Ofek and D. Yermack. 1997. Managerial Entrenchment and Capital Structure Decisions. *Journal of Finance* 52, 1411-1438.

Bhojraj, S., C. Lee and D. Oler. 2003. What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research* 41*,* 745-774.

Botosan, C. and M. Harris. 2000. Motivations for a Change in Disclosure Frequency and Its Consequences: An Examination of Voluntary Quarterly Segment Disclosures. *Journal of Accounting Research* 38*,* 329-353.

Botosan, C. and M. Stanford. 2005. Managers' Motives to Withhold Segment Disclosures and the Effect of SFAS No. 131 on Analysts' Information Environment. *The Accounting Review* 80: 751 – 771.

Denis, D., D. Denis, and A. Sarin. 1997. Agency Problems, Equity Ownership, and Corporate Diversification. *The Journal of Finance* 52, 135-160.

Ettredge, M., S. Kwon, D. Smith, and M. Stone. 2006. The Effect of SFAS No. 131 on the Cross-Segment Variability of Profits Reported by Multiple Segment Firms. *Review of Accounting Studies* 11, 91-117.

Financial Accounting Standards Board (FASB). 1976. *Financial Reporting for Segments of a Business Enterprise.* Statement of Financial Accounting Standards No. 14. Norwalk CT (FASB).

Financial Accounting Standards Board (FASB). 1997. *Disclosures about Segments of an Enterprise and Related Information*. Statement of Financial Accounting Standards No. 131. Norwalk CT (FASB).

Frankel, R., M. McNichols and G. Wilson. 1995. Discretionary Disclosure and External Financing. *The Accounting Review* 70, 135-150.

Grossman, S. 1981. The Role of Warranties and Private Disclosure about Product Quality. *Journal of Law and Economics* 24, 461-483.

Hall, B., C. Cummins, E. Laderman and J. Mundy. 1988. The R&D Master File Documentation. National Bureau of Economic Research, Boston, MA.

Harris, M. 1998. The Association between Competition and Managers' Business Segment Reporting Decisions. *Journal of Accounting Research* 36, 111-128.

Hayes, R. and R. Lundholm. 1996. Segment Reporting to the Capital Market in the Presence of a Competitor. *Journal of Accounting Research* 34, 261-279.

Hope, O. and W. Thomas. 2008. Managerial Empire Building and Firm Disclosure. *Journal of Accounting Research* 46, 591-626.

Lamont, O. 1997. Cash Flow and Investment: Evidence from Internal Capital Markets. *Journal of Finance* 52, 83-109.

Lang, M. and R. Lundholm. 1993. Cross-sectional Determinants of Analyst Ratings of Corporate Disclosures. *Journal of Accounting Research* 31, 246-271.

______________. 1996. Corporate Disclosure Policy and Analyst Behavior. *The Accounting Review* 71, 467-492.

Lang, L. and R. Stulz. 1994. Tobin's Q, Corporate Diversification and Firm Performance. *Journal of Political Economy* 102, 1248-1280.

Liberty, S. and J. Zimmerman. 1986. Labor Union Contract Negotiations and Accounting Choices. *The Accounting Review* 61, 692-712.

Lindenberg, E. and S. Ross. 1981. Tobin's q and industrial organization. *Journal of Business* 54, 1-32.

Maksimovic, V., G. Phillips and N. Prabhala. 2008. Post-Merger Restructuring and the Boundaries of the Firm, Working Paper, University of Maryland.

McGuckin, R. and G. Pascoe. 1988. The Longitudinal Research Database: Status and Research Possibilities. *Survey of Current Business* 68, 30-37.

Milgrom, P. 1981. Good News and Bad News: Representation Theorems and Applications. *Bell Journal of Economics* 12, 380-391.

Norton, E., H. Wang and C. Ai. 2004. Computing Interaction Effects and Standard Errors in Logit and Probit Models. *Stata Journal* 4, 103-116.

Powers, E. 2005. Interpreting Logit Regressions with Interaction Terms: An Application to the Management Turnover Literature. *Journal of Corporate Finance* 11, 504-522.

Rajan, R,, H. Servaes and L. Zingales. 2000. The Cost of Diversity: The Diversification Discount and Inefficient Investment. *Journal of Finance* 55, 35-80.

Sanders, J., S. Alexander and S. Clark. 1999. "New Segment Reporting: Is It Working?" *Strategic Finance* December, 35-38.

Scott, T. 1994. Incentives and Disincentives for Financial Disclosure: Voluntary Disclosure of Defined Benefit Pension Plan Information by Canadian Firms. *The Accounting Review* 69, 26-43.

Shin, H. and R. Stulz. 1998. Are Internal Capital Markets Efficient? *Quarterly Journal of Economics* 113, 531-552.

Tang, V. 2009. Proprietary Costs and Privately Held Firms' Financing Choice between Public Offerings and Private Placements. Working paper (Georgetown University).4

Verrecchia, R. 1983. Discretionary Disclosure. *Journal of Accounting and Economics* 5, 365-380.

Verrecchia, R. and J. Weber. 2006. Redacted Disclosure. *Journal of Accounting Research* 44, 791-814.

Table 1
Comparison of Final Sample to Compustat

Panel A – Reconciliation of Compustat Sample and Final Sample

		Firm-years
In Compustat		9,975
Less:	Not in the Census Compustat bridge file	4,195
	Has only one pseudo-segment	2,667
	Aggregate TVS per census exceeds 125% of SALES per Compustat	323
	Aggregate TVS per census is less than 75% of SALES per Compustat	1,165
Final sample		1,625

Panel B – Firm Level Descriptive Statistics

Pseudo-Segments			Compustat Segments		
Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
5.1	3	5.1	1.6	1	0.9

Panel C – Industry Composition

Two-digit SIC Code	Industry Description	Sample: Percentage of Pseudo-Segments	Sample: Percentage of TVS	Compustat Manufacturing: Percentage of Segments	Compustat Manufacturing: Percentage of Sales	All Compustat: Percentage of Segments	All Compustat: Percentage of Sales
20 & 21	Food & Tobacco Products	5.95	18.18	5.20	11.67	5.43	12.12
22	Textiles	3.63	3.06	1.84	0.94	1.85	0.92
23	Apparel & Other Finished Products	2.70	1.86	2.09	0.90	2.01	0.88
24	Lumber and Wood Products	3.55	2.29	1.75	1.27	1.79	1.25
25	Furniture and Fixtures	2.67	1.55	1.47	0.71	1.44	0.68
26	Paper and Allied Products	4.45	7.22	2.57	2.57	2.61	2.61
27	Printing & Publishing	4.42	5.92	2.83	1.38	3.69	2.03
28 & 29	Chemicals & Petroleum	5.73	6.88	14.46	20.41	14.30	20.54
30 & 31	Rubber, Plastics & Leather	8.06	3.08	4.27	1.81	4.29	1.82
32	Stone, Clay, Glass & Concrete	2.78	1.35	1.93	0.87	1.98	0.96
33	Primary Metals	5.76	6.42	3.85	3.84	3.87	3.81
34	Fabricated Metals & Transportation Equipment	11.34	4.73	4.86	1.51	5.00	1.55
35	Industrial/Commercial Machinery & Computers	13.33	6.26	16.65	13.28	16.38	12.81
36	Electrical Equipment	12.20	9.11	15.05	10.93	14.67	11.10
37	Transportation Equipment	5.35	16.55	5.35	22.57	5.28	21.53
38	Measuring Instruments, Photo Goods & Watches	6.42	4.71	13.29	4.77	12.88	4.82
39	Miscellaneous	1.67	0.84	2.54	0.58	2.55	0.58
Total		100.00	100.00	100.00	100.00	100.00	100.00

Panel A provides a reconciliation of all observations on Compustat for the years 1987, 1992, and 1997 and the observations that make up our sample of 1,625 firm-years (1,008 unique firms) for the years 1987, 1992 and 1997. A firm is included in our sample if its primary SIC code is in the manufacturing sector (SIC 2000-3999) and it can be matched from Compustat to the Longitudinal Research Database (LRD) of the U.S. Census Bureau. Panel B provides descriptive statistics for our sample. Panel C compares the industry composition of the pseudo-segments in our sample to the industry composition of Compustat line-of-business segments. In Panel C All Compustat refers to a sample of 13,530 segments (10,442 firm-years, 5,863 firms) that have non-negative sales and assets per the Compustat database for the years 1987, 1992, or 1997, and have a primary or secondary SIC code between 2000 and 3999. Compustat Manufacturing relates to 12,358 segments (9,585 firm-years, 5,441 firms) that are a subset of All Compustat and for which 75% of firm-level sales is attributable to manufacturing segments (i.e., segments with either a primary or secondary four-digit SIC code between 2000 and 3999).

A *pseudo-segment* is defined as all LRD plants of a firm that operate in the same four-digit SIC code. *TVS* denotes total value of shipments for a particular pseudo-segment per the LRD. In Panel A *Aggregate TVS* is defined as follows:

$$Aggregate\,TVS = \sum_{j=1}^{n} TVS_j$$

In the above equation j is a pseudo-segment index and n is the number of pseudo segments for a particular firm. *SALES* denotes annual sales per Compustat (i.e., data item 12). *Compustat segments* reflect line-of-business segments reported by the firm in its external reports filed with the SEC. In Panel C: percentage of pseudo-segments represents the percentage of all sample pseudo segments in a particular industry; percentage of *TVS* represents the percentage of the total value of shipments reported by all pseudo-segments that is attributable to the pseudo-segments of a particular industry; percentage of segments represents the percentage of all manufacturing segments in a particular industry; and, percentage of sales represents the percentage of all sales reported by manufacturing segments that is attributable to segments of a particular industry.

Table 2
Pseudo-Segment Level Descriptive Statistics

Panel A – All Sample Firms

	MEAN		MEDIAN		STANDARD DEVIATION	
	MATCH = 1	MATCH = 0	MATCH = 1	MATCH = 0	MATCH = 1	MATCH = 0
I_PROFIT	0.15	0.05 ***	0.11	0.08 ***	0.69	0.85
PROFITADJ	0.32	0.35 ***	0.29	0.34 ***	0.39	0.37
PRIVATE	0.56	0.63 ***	0.58	0.65 ***	0.20	0.20
TRANSIN	0.26	0.32 ***	0.00	0.00 ***	0.44	0.46
LOENT_HISUB	0.28	0.28	0.00	0.00	0.45	0.45
LABOR	0.21	0.22 ***	0.21	0.22 ***	0.07	0.08
FSIZE	5.56	6.26 ***	5.49	6.29 ***	1.66	1.73
FOURFIRM	0.34	0.30 ***	0.31	0.26 ***	0.18	0.18
RELSIZE	0.36	0.10 ***	0.26	0.04 ***	0.30	0.17
SEGDIVERSITY	0.58	0.49 ***	0.50	0.46 ***	0.24	0.23
NUMEST	7.04	11.95 ***	5.00	8.00 ***	6.65	10.91
CEN_CMPSTAT	0.96	0.95 ***	0.96	0.94 ***	0.12	0.12
INDMATCHRATE	0.42	0.32 ***	0.41	0.33 ***	0.26	0.24
NEC	0.20	0.25 ***	0.00	0.00 ***	0.40	0.44
N	3,032	5,255				

Table B – Single-Segment Firms

	MEAN		MEDIAN		STANDARD DEVIATION	
	MATCH = 1	MATCH = 0	MATCH = 1	MATCH = 0	MATCH = 1	MATCH = 0
I_PROFIT	0.14	0.05 ***	0.11	0.07 ***	0.73	0.87
PROFITADJ	0.34	0.34	0.30	0.34	0.39	0.38
PRIVATE	0.55	0.63 ***	0.56	0.66 ***	0.20	0.20
TRANSIN	0.25	0.31 ***	0.00	0.00 ***	0.43	0.46
LOENT_HISUB	0.27	0.28	0.00	0.00	0.44	0.45
LABOR	0.20	0.21 ***	0.20	0.21 ***	0.07	0.08
FSIZE	4.99	5.61 ***	4.89	5.54 ***	1.43	1.65
FOURFIRM	0.34	0.29 ***	0.31	0.26 ***	0.18	0.17
RELSIZE	0.53	0.16 ***	0.55	0.07 ***	0.31	0.21
SEGDIVERSITY	0.63	0.55 ***	0.50	0.50 ***	0.25	0.25
NUMEST	3.51	5.97 ***	3.00	4.00 ***	2.15	4.69
CEN_CMPSTAT	0.97	0.96 **	0.98	0.97 **	0.12	0.12
INDMATCHRATE	0.42	0.32 ***	0.41	0.33 ***	0.25	0.25
NEC	0.19	0.25 ***	0.00	0.00 ***	0.40	0.43
N	1,175	2,125				

Panel C – Multisegment Firms

	MEAN			MEDIAN			STANDARD DEVIATION	
	MATCH = 1	MATCH = 0		MATCH = 1	MATCH = 0		MATCH = 1	MATCH = 0
I_PROFIT	0.15	0.05	***	0.10	0.08	***	0.67	0.84
PROFITADJ	0.30	0.35	***	0.28	0.34	***	0.39	0.36
PRIVATE	0.56	0.62	***	0.59	0.65	***	0.21	0.20
TRANSIN	0.27	0.32	***	0.00	0.00	***	0.44	0.47
LOENT_HISUB	0.29	0.28		0.00	0.00		0.45	0.45
LABOR	0.21	0.22	***	0.21	0.22	***	0.07	0.08
FSIZE	5.92	6.71	***	5.92	6.82	***	1.69	1.64
FOURFIRM	0.34	0.30	***	0.31	0.26	***	0.19	0.19
RELSIZE	0.25	0.06	***	0.17	0.02	***	0.24	0.11
SEGDIVERSITY	0.55	0.45	***	0.50	0.40	***	0.23	0.20
NUMEST	9.27	16.00	***	7.00	13.00	***	7.51	12.01
CEN_CMPSTAT	0.94	0.93	***	0.94	0.92	***	0.12	0.12
INDMATCHRATE	0.42	0.33	***	0.41	0.33	***	0.26	0.24
NEC	0.20	0.26	***	0.00	0.00	***	0.40	0.44
N	1,857	3,130						

Panel A shows descriptive statistics for 8,287 pseudo-segments of 1,625 firm-years (1,008 unique firms) over the years 1987, 1992, and 1997. Panel B shows descriptive statistics for 3,300 pseudo-segments of 956 firm-years over the years 1987, 1992, and 1997. Panel C shows descriptive statistics for 4,987 pseudo-segments of 669 firm-years over the years 1987, 1992, and 1997. The pseudo-segments described in Panel B all relate to firms that have one line-of-business segment per Compustat (i.e., single-segment firms). The pseudo-segments described in Panel C all relate to firms that have multiple line-of-business segments per Compustat (i.e., multisegment firms). Please refer to Table 1 for additional sample selection criteria.

A *pseudo-segment* is defined as all LRD plants of a firm that operate in the same four-digit SIC code. *MATCH* equals one if the pseudo-segment four-digit SIC code matches the primary or secondary SIC code of a Compustat segment, zero otherwise. *I_PROFIT* equals the difference between the pseudo-segment's gross margin and the industry average gross margin, divided by the standard deviation of gross margin across the industry; gross margin is obtained from the LRD database, and equals total value of shipments (*TVS*) less cost of materials less salaries and wages, with the difference scaled by *TVS*. In a regression estimated from all firms in the pseudo-segment's industry where the dependent variable is current year gross margin and the two independent variables are lagged gross margin if negative, and lagged gross margin if positive, the coefficient from the positive gross margin realizations is defined as *PROFITADJ*. *PRIVATE* equals the ratio of the sum of *TVS* across all firms in the pseudo-segment's industry that cannot be linked to Compustat (i.e., we assume that firms that are not on Compustat are privately held), divided by the sum of *TVS* across all firms in the industry. The variable *IFAIR* is used to calculate *TRANSIN*. *IFAIR* equals the proportion of a pseudo-segment's establishments that were not owned by the firm in the previous census year less the proportion of new plants for the entire industry (excluding the firm) less the weighted (by *TVS*) firm average of this difference across all of its pseudo-segments. In equation form:

$$IFAIR_j = \frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} - \sum_{j=1}^{n} w_j \left(\frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} \right)$$

In the equation, *NP* denotes new plants, *EP* ending plants, *w* is *RELSIZE* for the pseudo-segment, *j* indexes pseudo-segment, and *i* indexes the industry. *TRANSIN* equals one if *IFAIR* is positive for the pseudo-segment and the industry Tobin's Q for that pseudo-segment is less than the average of all other industry Tobin's Q measures for the firm, zero otherwise. *LOENT_HISUB* is an indicator variable that equals either zero or one. *LOENT_HISUB* equals one if the pseudo-segment is a member of an industry for which: (1) the ratio of industry-level capital spending to industry-level TVS in year t is below the year t median of all manufacturing industries and (2) the ratio of industry-level TVS to the sum of industry-level raw materials costs and industry-level payroll costs for year t is above the year t median of all manufacturing industries. *LABOR* equals the ratio of total labor costs incurred by establishments in the same industry as the pseudo-segment to the sum of *TVS* for all establishments in the same industry as the pseudo-segment. *FOURFIRM* equals the sum of *TVS* for the four largest firms in the same industry as the pseudo-segment divided by the sum of *TVS* for all firms in the same industry as the pseudo-segment. *RELSIZE* equals the *TVS* of the pseudo-segment deflated by the total *TVS* summed across all of the firm's pseudo-segments. *SEGDIVERSITY* equals number of unique two-digit SIC codes across the firm's pseudo-segments deflated by the total number of pseudo-segments. *CEN_CMPSTAT* equals total *TVS* summed across all of the firm's pseudo-segments deflated by total sales of the firm per Compustat. *INDMATCHRATE* equals the proportion of pseudo-segments in an industry where *MATCH* equals one, excluding the sample pseudo-segment being analyzed. *NEC* equals one if the four digit SIC code contains the phrase "Not Elsewhere Classified" in the SIC manual. *NUMEST* equals the number of pseudo-segments operated by the firm. N denotes the number of observations. ***, **, and * denote significance at the 1%, 5% and 10% levels.

Table 3
Correlation Matrix for Multisegment Firms

	MATCH	I_PROFIT	PROFITADJ	PRIVATE	TRANSIN	LOENT_HISUB	LABOR	FSIZE	FOURFIRM	RELSIZE	SEGDIVERSITY	NUMEST	CEN_CMPSTAT	INDMATCHRATE	NEC
MATCH		0.064	-0.064	-0.142	-0.054	0.011	-0.085	-0.224	0.093	0.465	0.232	-0.295	0.047	0.182	-0.064
		<.0001	<.0001	<.0001	0.000	0.432	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.001	<.0001	<.0001
I_PROFIT	0.056		0.029	-0.017	0.003	-0.009	-0.008	-0.031	0.016	0.072	0.007	0.005	0.001	-0.001	-0.030
	<.0001		0.040	0.235	0.849	0.526	0.557	0.028	0.254	<.0001	0.616	0.704	0.924	0.949	0.032
PROFITADJ	-0.049	0.015		0.185	-0.016	-0.072	-0.068	0.032	-0.016	-0.039	-0.004	0.000	-0.006	-0.096	-0.014
	0.001	0.277		<.0001	0.261	<.0001	<.0001	0.025	0.274	0.006	0.766	0.995	0.683	<.0001	0.329
PRIVATE	-0.139	-0.024	0.152		0.038	0.109	0.175	-0.133	-0.705	-0.192	0.041	-0.038	0.010	-0.310	0.108
	<.0001	0.090	<.0001		0.008	<.0001	<.0001	<.0001	<.0001	<.0001	0.003	0.008	0.469	<.0001	<.0001
TRANSIN	-0.054	0.003	-0.008	0.040		0.065	0.020	0.052	0.012	-0.052	-0.028	0.066	-0.035	0.022	-0.087
	0.000	0.845	0.567	0.005		<.0001	0.154	0.000	0.389	0.000	0.050	<.0001	0.015	0.123	<.0001
LOENT_HISUB	0.011	-0.022	-0.048	0.117	0.065		-0.139	-0.005	0.001	0.014	-0.013	0.012	0.042	0.048	-0.133
	0.432	0.118	0.001	<.0001	<.0001		<.0001	0.708	0.957	0.311	0.363	0.382	0.003	0.001	<.0001
LABOR	-0.083	0.009	-0.116	0.202	0.026	-0.134		-0.172	-0.290	-0.069	0.115	0.001	-0.042	-0.119	0.097
	<.0001	0.533	<.0001	<.0001	0.066	<.0001		<.0001	<.0001	<.0001	<.0001	0.938	0.003	<.0001	<.0001
FSIZE	-0.225	-0.027	0.037	-0.134	0.050	-0.002	-0.160		0.172	-0.458	-0.616	0.567	-0.228	0.018	-0.066
	<.0001	0.061	0.008	<.0001	0.000	0.889	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	0.194	<.0001
FOURFIRM	0.099	0.021	0.024	-0.687	0.018	0.012	-0.252	0.177		0.102	-0.082	0.063	-0.016	0.176	-0.279
	<.0001	0.136	0.084	<.0001	0.208	0.413	<.0001	<.0001		<.0001	<.0001	<.0001	0.273	<.0001	<.0001
RELSIZE	0.536	0.108	-0.037	-0.177	-0.046	0.009	-0.066	-0.379	0.118		0.383	-0.381	0.075	0.161	-0.065
	<.0001	<.0001	0.009	<.0001	0.001	0.530	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001
SEGDIVERSITY	0.219	0.008	-0.006	0.028	-0.030	-0.013	0.097	-0.616	-0.097	0.383		-0.591	0.072	0.033	0.077
	<.0001	0.579	0.670	0.049	0.037	0.343	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	0.019	<.0001
NUMEST	-0.342	0.008	0.022	-0.011	0.049	-0.004	0.035	0.675	0.052	-0.540	-0.654		-0.170	-0.049	-0.007
	<.0001	0.573	0.128	0.418	0.001	0.752	0.014	<.0001	0.000	<.0001	<.0001		<.0001	0.001	0.639
CEN_CMPSTAT	0.051	-0.010	-0.007	0.008	-0.034	0.041	-0.038	-0.224	-0.020	0.087	0.058	-0.167		-0.014	-0.031
	0.000	0.460	0.640	0.563	0.015	0.004	0.007	<.0001	0.166	<.0001	<.0001	<.0001		0.318	0.028
INDMATCHRATE	0.186	0.011	-0.102	-0.327	0.023	0.046	-0.140	0.014	0.182	0.148	0.031	-0.028	-0.010		-0.098
	<.0001	0.422	<.0001	<.0001	0.105	0.001	<.0001	0.319	<.0001	<.0001	0.029	0.052	0.486		<.0001
NEC	-0.064	-0.015	-0.044	0.114	-0.087	-0.133	0.091	-0.061	-0.314	-0.058	0.079	-0.011	-0.028	-0.107	
	<.0001	0.294	0.002	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.433	0.046	<.0001	

Correlations are estimated using 4,987 pseudo-segments from the years 1987, 1992 and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. Pearson (Spearman) correlations are above (below) the diagonal. P-values for two-tailed tests are shown in the cell below the cell that contains the estimated correlation coefficient. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

Table 4
Logistic Regression Estimates for Pseudo-Segment Disclosure Decision

Panel A – All Sample Firms

	Predicted Sign	Coefficient	P-value	Economic Magnitude
INTERCEPT		-0.773	0.092	
Treatment Variables				
I_PROFIT	?	0.080	0.033	0.018
PROFITADJ	-	-0.139	0.128	-0.032
PRIVATE	-	-0.471	0.035	-0.108
TRANSIN	-	-0.157	0.004	-0.036
LOENT_HISUB	-	0.050	0.493	0.011
LABOR	-	-0.675	0.171	-0.155
Control Variables				
FSIZE		0.015	0.580	0.003
FOURFIRM		0.214	0.408	0.049
RELSIZE		4.073	0.000	0.933
SEGDIVERSITY		0.039	0.838	-0.093
NUMEST		-0.025	0.000	-0.006
CEN_CMPSTAT		-0.407	0.178	-0.093
INDMATCHRATE		1.014	0.000	0.232
NEC		-0.107	0.193	-0.024
Pseudo R-squared		0.205		

The regression is estimated using 8,287 pseudo-segments from the years 1987, 1992 and 1997. All p-values are two-tailed. The *Economic Magnitude* column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals one) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

Panel B – Single-Segment Firms

	Predicted Sign	Coefficient	P-value	Economic Magnitude
INTERCEPT		-1.040	0.078	
Treatment Variables				
I_PROFIT	?	-0.020	0.716	-0.004
PROFITADJ	-	0.134	0.362	0.029
PRIVATE	-	-0.638	0.053	-0.140
TRANSIN	-	-0.061	-0.548	-0.013
LOENT_HISUB	-	0.086	0.480	0.019
LABOR	-	-0.904	-0.263	-0.198
Control Variables				
FSIZE		-0.091	0.023	-0.020
FOURFIRM		0.375	0.350	0.082
RELSIZE		4.313	0.000	0.946
CEN_CMPSTAT		-0.325	0.404	-0.071
INDMATCHRATE		0.779	0.000	0.171
NEC		-0.035	0.800	-0.008
Pseudo R-squared		0.285		

The regression is estimated using 3,300 pseudo-segments from the years 1987, 1992 and 1997. The pseudo-segments all relate to firms that have a single line-of-business segment per Compustat. The binary dependent variable is *MATCH*. All p-values are two-tailed. The *Economic Magnitude* column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals one) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

Panel C – Multisegment Firms

	Predicted Sign	Coefficient	P-value	Economic Magnitude
INTERCEPT		-0.480	0.430	
Treatment Variables				
I_PROFIT	?	0.122	0.020	0.028
PROFITADJ	-	-0.314	0.005	-0.073
PRIVATE	-	-0.226	0.404	-0.052
TRANSIN	-	-0.182	0.009	-0.042
LOENT_HISUB	-	-0.034	0.712	-0.008
LABOR	-	-1.363	0.017	-0.316
Control Variables				
FSIZE		0.020	0.574	0.005
FOURFIRM		0.229	0.450	0.053
RELSIZE		5.816	0.000	1.349
SEGDIVERSITY		-0.053	0.838	-0.012
NUMEST		-0.045	0.000	-0.010
CEN_CMPSTAT		-0.207	0.550	-0.048
INDMATCHRATE		1.089	0.000	0.253
NEC		-0.141	0.169	-0.032
Pseudo R-squared		0.219		

The regression is estimated using 4,987 pseudo-segments from the years 1987, 1992 and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. The binary dependent variable is *MATCH*. All p-values are two-tailed. The *Economic Magnitude* column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals one) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

Table 5
Logistic Regression Estimates for Pseudo-Segment Disclosure Decision of Multisegment Firms
Interactions between Treatment Variables and *SML_RELSIZE* are Included

	Predicted Sign	Coefficient	P-value	Economic Magnitude
INTERCEPT		0.605	0.347	
SML_RELSIZE		-1.558	0.000	-0.359
Treatment Variables				
I_PROFIT	?	0.073	0.489	0.017
I_PROFIT×SML_RELSIZE	?	0.051	0.676	0.012
PROFITADJ	-	-0.459	0.020	-0.105
PROFITADJ×SML_RELSIZE	?	0.230	0.323	0.052
PRIVATE	-	-0.202	0.560	-0.046
PRIVATE×SML_RELSIZE	?	0.107	0.810	0.024
TRANSIN	-	-0.238	0.042	-0.053
TRANSIN×SML_RELSIZE	?	0.089	0.558	0.020
LOENT_HISUB	-	0.025	0.868	0.006
LOENT_HISUB×SML_RELSIZE	?	-0.120	0.512	-0.027
LABOR	-	-2.723	0.005	-0.620
LABOR×SML_RELSIZE	?	2.109	0.061	0.480
Control Variables				
FSIZE		0.035	0.346	0.008
FOURFIRM		0.245	0.418	0.056
RELSIZE		3.350	0.000	0.763
SEGDIVERSITY		-0.008	0.975	-0.002
NUMEST		-0.043	0.000	-0.010
CEN_CMPSTAT		-0.224	0.524	-0.051
INDMATCHRATE		1.083	0.000	0.247
NEC		-0.140	0.181	-0.031
Pseudo R-squared		0.233		

The regression is estimated using 4,987 pseudo-segments from the years 1987, 1992 and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. The binary dependent variable is *MATCH*. *SML_RELSIZE* is an indicator variable that equals one (zero) if *RELSIZE* is (not) less than 0.10. All p-values are two-tailed. The *Economic Magnitude* column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals one) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

Table 6
Logistic Regression Estimates for Pseudo-Segment Disclosure Decision of Multisegment Firms Separate Regressions Estimated for Firms with Negative and Nonnegative Excess Values from Diversification

Panel A – Excess Value from Diversification is Negative

	Predicted Sign	Coefficient	P-value	Economic Magnitude
INTERCEPT		-0.375	0.567	
Treatment Variables				
I_PROFIT	?	0.153	0.028	0.037
PROFITADJ	-	-0.229	0.135	-0.055
PRIVATE	-	-0.302	0.373	-0.072
TRANSIN	-	-0.175	0.091	-0.041
LOENT_HISUB	-	-0.009	0.936	-0.002
LABOR	-	-2.360	0.001	-0.563
Control Variables				
FSIZE		0.030	0.425	0.007
FOURFIRM		0.567	0.126	0.135
RELSIZE		5.142	0.000	1.226
SEGDIVERSITY		0.299	0.316	0.071
NUMEST		-0.041	0.000	-0.010
CEN_CMPSTAT		-0.402	0.290	-0.096
INDMATCHRATE		1.056	0.000	0.252
NEC		-0.194	0.152	-0.046
Pseudo R-squared		0.204		

Panel B – Excess Value from Diversification is Nonnegative

	Predicted Sign	Coefficient	P-value	Economic Magnitude
INTERCEPT		-2.126	0.048	
Treatment Variables				
I_PROFIT	?	0.117	0.124	0.026
PROFITADJ	-	-0.399	0.007	-0.088
PRIVATE	-	0.113	0.794	0.025
TRANSIN	-	-0.201	0.087	-0.044
LOENT_HISUB	-	-0.026	0.873	-0.006
LABOR	-	0.313	0.733	0.069
Control Variables				
FSIZE		0.065	0.332	0.014
FOURFIRM		-0.091	0.858	-0.020
RELSIZE		7.028	0.000	1.555
SEGDIVERSITY		-0.283	0.531	-0.063
NUMEST		-0.046	0.000	-0.010
CEN_CMPSTAT		0.678	0.239	0.150
INDMATCHRATE		1.266	0.000	0.280
NEC		-0.042	0.779	-0.009
Pseudo R-squared		0.239		

The regressions are estimated using 4,517 pseudo-segments from the years 1987, 1992 and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. The 2,545 (1,972) pseudo-segments underlying the regression in Panel A (B) relate to firm-years in which the excess value of diversification as measured using the sales-multiple approach discussed in Berger and Ofek (1995) is negative (non-negative). The binary dependent variable is *MATCH*. The *Economic Magnitude* column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals one) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.