THE JOURNAL OF FINANCE • VOL. LIX, NO. 2 • APRIL 2004

Diversification Discount or Premium? New Evidence from the Business Information Tracking Series

BELÉN VILLALONGA*

ABSTRACT

I use the Business Information Tracking Series (BITS), a new census database that covers the whole U.S. economy at the establishment level, to examine whether the finding of a diversification discount is an artifact of segment data. BITS data allow me to construct business units that are more consistently and objectively defined than segments, and thus more comparable across firms. Using these data on a sample that yields a discount according to segment data, I find a diversification premium. The premium is robust to variations in the sample, business unit definition, and measures of excess value and diversification.

THE DIVERSIFICATION DISCOUNT has been the subject of an active debate in corporate finance during the past few years.[1] On the one hand, Lang and Stulz (1994), Berger and Ofek (1995), and Servaes (1996) find that diversified firms trade at an average discount relative to single-segment firms. This finding has often been interpreted as evidence that diversification destroys value.[2] On the other hand, several studies show that the discount is only the product of sample selection biases. Villalonga (1999) and Campa and Kedia (2002) find that diversified firms traded at a discount prior to diversifying. More generally, diversifying and nondiversifying firms differ systematically in multiple characteristics. When the selection bias is corrected for, the diversification discount

*Belén Villalonga is at Harvard University. I would like to thank Judith Chevalier, Harold Demsetz, Amy Dittmar, Connie Helfat, Armen Hovakimian, Guido Imbens, Matthias Kahl, Walter Kuemmerle, Vojislav Maksimovic, Richard Makadok, Anita McGahan, Bill McKelvey, William Ouchi, Thomas Piper, Erik Powers, David Scharfstein, Olav Sorenson, Richard Roll, Richard Ruback, Richard Rumelt, the editor (Richard Green), an anonymous referee, and seminar participants at Harvard, MIT, and the University of Michigan for their comments and suggestions. Thanks to Nelson Lim and Jim Davis for their help at the California and Boston Census Research Data Centers (CCRDC and BRDC), respectively. Financial support from the Division of Research at the Harvard Business School and the CCRDC is also gratefully acknowledged. The research in this paper was conducted while the author was a research associate at the CCRDC and at the BRDC in association with the Center for Economic Studies of the U.S. Bureau of the Census. Research results and conclusions expressed are those of the author and do not necessarily indicate concurrence by the Bureau of the Census, the Center for Economic Studies, the CCRDC, or the BRDC. All errors remain mine.

[1] See Martin and Sayrak (2003) for a complete review of the literature.

[2] For instance, Berger and Ofek in a later paper claim that "[Berger and Ofek (1995)] find that, during 1986–1991, the average diversified firm destroyed about 15 percent of the value its lines of business would have had if operated as stand-alone businesses" (1996, p. 1175).

disappears or even turns into a premium. Graham, Lemmon, and Wolf (2002) show that half or more of the discount appears because the segments acquired by diversifying firms were also discounted prior to their acquisition. Given that both diversifying firms and their targets trade at a discount prior to diversification, it is not surprising that diversified firms exhibit a discount. However, the findings of these studies suggest that diversification, in itself, does not destroy value.

This paper takes the skepticism toward the diversification discount one step further by questioning the finding itself, not just its interpretation as evidence of value destruction. Specifically, I investigate the possibility that the discount is an artifact of segment data. COMPUSTAT provides disaggregated financial information for business segments that represent at least 10 percent of a firm's sales, assets, or profits. Prior studies of the diversification discount have used this information to measure corporate diversification in U.S. firms. The data thus determine which firms are diversified and which are not, and the industry or industries in which each firm operates. This, in turn, affects the estimated valuation effects by determining which market values to impute to a firm's "pieces" in order to compare the sum of the pieces to the firm's market value as a whole.

To assess the effect of corporate diversification on firm value, it is therefore crucial to measure diversification correctly in the first place. Yet, the use of segment data for this purpose raises several concerns. First, the extent of disaggregation in segment financial reporting is much lower than the true extent of a firm's industrial diversification (Lichtenberg (1991)). Firms are required by the Financial Accounting Standards Board (FASB) to report disaggregated information for segments that meet the 10 percent materiality condition. Effectively, this means that the maximum number of different industries that can be observed for any given firm is 10. Because managers have considerable discretion in disclosing segment-level information, the actual number of segments reported by some firms seems to be even lower.[3] In contrast, sources that do not impose such censoring on the data show that the number of industries in which a firm is present can be much higher. This is particularly the case when industries are defined at the four-digit SIC code level of precision, as they typically are in COMPUSTAT. For instance, the maximum number of four-digit SIC codes for a single firm in my sample is 133.[4,5]

[3] Segment under-reporting was one of the major concerns that triggered the issuance of the Statement of Financial Accounting Standards (SFAS) 131 in 1997 (Association for Investment Management Research (AIMR) (1993); American Institute of Certified Public Accountants (AICPA) (1994)). SFAS 131 superseded SFAS 14 in the regulation of segment reporting (FASB (1997)). As expected, its implementation has resulted in a greater number of segments being reported by at least certain firms (Berger and Hann (2002, 2003)). However, these improvements do not affect diversification discount studies, whose samples are all pre-1997.

[4] The name of the company or companies with 133 SIC codes cannot be reported because the Census Bureau's confidentiality policy prohibits the disclosure of any information about individual firms.

[5] The percentage of firms present in more than 10 industries is as high as 56 percent for the 500 largest U.S. public companies (Montgomery (1994), or 17% for all COMPUSTAT firms (Lichtenberg (1991)).

A second concern arises from the definition of segment itself. The Statement of Financial Accounting Standards (SFAS) 14 defines a segment as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit" (FASB (1976), paragraph 10a). Hence, segments, by definition, can be an aggregation of two or more activities, vertically or otherwise related.[6] The aggregation of activities into any given segment SIC code differs from firm to firm. This difference is exacerbated by the fact that segments are self-reported: Segments are identified by name by the reporting company, and assigned a four-digit SIC code by COMPUSTAT. Davis and Duhaime (1992) find that, in 5–10 percent of cases, firms had grouped into one-segment businesses that were totally unrelated. A question therefore arises about the comparability of segments across firms.

A third concern is that firms frequently change the segments they report when there is no real change in their operations. Denis, Denis, and Sarin (1997) and Hyland (1997) find that about one-fourth of all changes in firms' number of segments in COMPUSTAT are purely reporting changes, as opposed to real instances of diversification or refocusing. The inconsistency in segment definitions occurs not only across firms, but also within firms over time.

The three concerns identified suggest that the use of segment data introduces bias at several points in the estimation of diversification's value effect. Firms may be misallocated to industries, and industries may be misallocated to firms. Firms that are present in more than one industry may get misclassified as nondiversified. This misclassification may in turn distort the industry mean or median qs that serve as benchmarks for the valuation of segments. Only single-business firms are supposed to enter the computation, but in fact some of the single-segment firms included are multibusiness. In other words, the "pure-play" qs imputed to segments may not be so pure. As a result of all this, estimates of the diversification discount based on segment data may be very different from those that can be obtained from other data sources. Whether segment data will yield a higher or lower discount than other sources is ultimately an empirical question. This is the central question investigated in this paper.

I use the Business Information Tracking Series (BITS) as an alternative data source to estimate the value effect of diversification. BITS is a new census database that covers the whole U.S. economy at the establishment level. These data allow me to construct business units that overcome the problems of segments described earlier. I use a common sample of firms and a common method (Lang and Stulz's) to compare the value estimates obtained on BITS against those obtained on COMPUSTAT. Consistent with earlier studies, I find a diversification discount when firms' activities are broken down into COMPUSTAT segments. When the same firms' activities are broken down into BITS business units, however, diversified firms trade at a significant average *premium* relative to comparable portfolios of single-business firms.

[6] As Davis and Duhaime put it, "either vertical integration *or* relatedness are necessary conditions for assigning two businesses to a single segment; vertical integration is also a sufficient condition for assigning two businesses to a single segment, but relatedness is not" (1992, p. 512).

After ruling out several candidate explanations for this result, I offer two plausible explanations: "Relatedness" and "strategic accounting." The relatedness explanation suggests that the two databases provide different but complementary measurements of diversification. Specifically, the findings in this paper can be interpreted as evidence that there is a discount to unrelated (conglomerate) diversification, but a premium to related diversification. Because related diversification is likely to predominate over conglomeration, when all diversification types are pooled together as they are in BITS, the net effect on firm value is a premium. The strategic accounting explanation suggests that the discount arises in segment data because diversified firms aggregate their activities into segments in ways that may make them appear as artificially low performers relative to single-segment firms in the same industries.

The two explanations are not mutually exclusive, and each of them is consistent with empirical evidence in other studies (e.g., Rumelt (1974) and Berger and Ofek (1995) for relatedness, and Harris (1998) and Berger and Hann (2002, 2003) for strategic accounting). A comparison between the segment SIC codes of single-segment firms in the sample and the SIC codes those same firms have in BITS does not enable me to eliminate either explanation. I therefore conclude that both relatedness and strategic accounting are likely to be responsible for the observed differences in results between BITS and COMPUSTAT data.

Two earlier studies have used establishment-level data from the U.S. Census Bureau to investigate issues of corporate diversification: Maksimovic and Phillips (2002) and Schoar (2002). Both papers use the Longitudinal Research Database (LRD) and focus on productivity as a measure of corporate performance. Schoar also looks at firm value using LRD-based measures of diversification and finds a discount, albeit smaller than that obtained using segment data on the same sample.

Two key differences between the analysis in this paper and Schoar's can account for the difference in our findings. First, Schoar uses COMPUSTAT-based industry qs rather than LRD-based qs. In this paper, I provide evidence that BITS-based industry qs are better measures than COMPUSTAT qs of the "true" industry average. Second, and perhaps more importantly, the LRD only covers manufacturing establishments (i.e., plants), whereas BITS also covers nonmanufacturing establishments. The difference in coverage is not trivial: Less than 20 percent of all multisegment firms included in COMPUSTAT's segment files are manufacturing-only; 56 percent are nonmanufacturing-only, and 24 percent are diversified across both sectors. Over 60 percent of the cross-sector diversifiers have less than 50 percent of their assets in manufacturing. When I restrict my sample to pure manufacturing firms to make it more comparable to Schoar's, I find a statistically insignificant premium. My findings thus confirm the importance of covering all sectors of the economy in a study of diversification.

The rest of the paper is organized as follows. Section I describes the data and measures of excess value and industry qs. Section II presents the results of the main analysis and several robustness checks. Section III discusses possible explanations for the results. Section IV concludes.

I. Data and Variables

A. Data and Sample

The sample in this study comes from two data sources: BITS and COMPUSTAT annual company and segment files. Both active and subsequently delisted COMPUSTAT companies are included. BITS and COMPUSTAT industry segment files are used to define the sample and to determine the proportion of each firm's activities by industry. COMPUSTAT geographic segment files are used to determine the percentage of U.S. versus non-U.S. operations for each firm. COMPUSTAT company-level files are used to obtain data on market values and other variables.

BITS provides establishment-level panel data between 1989 and 1996 for all U.S. private-sector establishments with positive payroll in any of those years, from both public and private firms. It includes a total of 50,708,528 establishment-year observations from 41,203,605 different firm-years. The database has been constructed by the U.S. Bureau of the Census under contract to the Office of Advocacy of the U.S. Small Business Administration, and is documented in Robb (2000). This paper is the first one to use these data.[7]

The basic unit of analysis in BITS is the business establishment, defined as "a single physical location where business is conducted or where services or industrial operations are performed." For each establishment-year observation, BITS contains information on its employment, annual payroll, primary four-digit SIC code, location, start year, the firm and legal entity to which the establishment belongs, and the firm's total employment. Establishments are owned by legal entities—typically corporations, partnerships, or sole proprietorships. A firm in BITS is defined as "the largest aggregation of business legal entities under common ownership and control."[8] Hence, firms may be composed of one or more legal entities, each of which may in turn own one or more establishments.

BITS provides the most complete coverage ever offered by an intra-firm level database. It contains data on the entire population of U.S. establishments from all sectors of the economy, excluding farms (SIC 01–02), railroads (SIC 40), the Postal Service (SIC 43), private households (SIC 88), and large pension, health, and welfare funds (SIC 6371 with at least 100 employees). There are at least three reasons why BITS can be considered the best source of data that is currently available to study corporate diversification in the United States. One, the establishment level of disaggregation allows for a breakdown of firms' activities by industry in a consistent way across firms. Two, establishments are linked longitudinally. Three, BITS covers all sectors of the economy. No other

[7] A preliminary version of BITS, which covered only 3 years and a slightly different set of variables and was then called the Longitudinal Enterprise and Establishment Microdata (LEEM), has been documented by Acs and Armington (1998) and used in Armington (1998), Armington and Robb (1998, 1999), and Acs, Armington, and Robb (1999).

[8] For the Census Bureau, a firm (Firm A) owns another firm (Firm B) if either of two basic criteria are met: (1) Firm A owns more than 50 percent of the voting stock of Firm B, or (2) Firm A has the power to direct the management and policies of Firm B (Nguyen (1998)).

data source offers all of these three features (Villalonga (2000)).[9] One important limitation of BITS, however, is that it contains no performance data, in particular the stock market values required to study the diversification discount. This paper overcomes the limitation using a merged BITS–COMPUSTAT sample. The data merging process is described in the Appendix.

Because COMPUSTAT covers all firms trading in U.S. stock markets, the firms in the sample are, by intersection, firms that were publicly traded in a U.S. stock exchange and had at least one establishment in the United States in any year between 1989 and 1996. As in prior studies of the discount, COMPUSTAT firms are only considered if they appear in both the company and the segment-level files and are not in the financial sector (SIC codes between 6000 and 6999). The sample also excludes establishments in agriculture (SIC codes lower than 1000), government (SIC 9000), and other noneconomic activities, such as membership organizations (SIC 8600), private households (SIC 8800), and unclassified services (SIC 8900). Also eliminated from the sample are firms with missing data for any of the required variables, and outliers. Following prior studies, firms are considered outliers if their imputed q is higher than four times (or lower than a fourth of) their true q.

The resulting BITS–COMPUSTAT Common Sample used in the main analysis is composed of 1,555,371 establishment-year observations from 12,708 different firm-years. For this study, all of a firm's establishments with a common SIC code have been aggregated into "business units." This is the BITS-based unit of analysis equivalent to COMPUSTAT's segments, in the sense of representing a firm's activities in an industry. Unlike segments, business units are constructed in a way that makes them comparable across firms, and the maximum number of business units for a firm is not limited to 10. The maximum number of business units within a firm is 133. On average, there are 122 establishments per firm: 15.5 establishments per business unit and 7.9 business units per firm. In contrast, the average number of segments per firm in the sample is 1.7, or 4.6 business units per segment.

Table I reports the number of firms, segments, business units, and establishments included in the Common Sample along with those in BITS and COMPUSTAT during the same period (1989–1996). This information can be used to compare the coverage and extent of disaggregation of firms' activities across the three databases. The coverage of BITS is not just different but also much greater than that of COMPUSTAT, which is in turn greater than that of the Common Sample. Nonetheless, the latter covers a substantive fraction of the overall population of COMPUSTAT firms (43 percent for all years, and as high as 53 percent in 1992). The number of BITS business units in the sample, which is nearly five times as large as the number of segments in the same firms, indicates a large information gain from combining the two data sources.

[9] Villalonga (2000) compares BITS with other large-sample databases that have been or may be used for academic research within firms. These include COMPUSTAT segment-level data, the FTC Line-Of-Business data, the Strategic Planning Institute's PIMS data, the U.S. Census Bureau's Longitudinal Research Database (LRD), and Trinet's Large Establishment Database.

Table I

Number of Firms, Segments, Business Units, and Establishments in the Common Sample, BITS, and COMPUSTAT

BITS refers to the *Business Information Tracking Series* of the U.S. Census Bureau. BITS covers all U.S. private-sector establishments with positive payroll in any year between 1989 and 1996, from both public and private firms. COMPUSTAT covers all firms that are publicly traded in U.S. stock markets. The figures reported for COMPUSTAT refer to all those active and inactive ("research") firms that appear in both the company and the segment COMPUSTAT files, and for which market values can be computed. The BITS–COMPUSTAT Common Sample has been constructed by merging both databases as described in the Appendix. An *establishment* is defined in BITS as "a single physical location where business is conducted or where services or industrial operations are performed," and a *firm* as "the largest aggregation of business legal entities (which are the legal owners of establishments) under common ownership and control." A *segment* is defined in SFAS 14 as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit." A *business unit* is defined in this paper as the aggregation of all of a firm's establishments with a common four-digit SIC code. Coverage ratios (i.e., the fraction of BITS or COMPUSTAT firms covered in the Common Sample) are in parentheses.

	Common Sample				BITS			COMPUSTAT	
Year	Establishments	Business Units	Segments	Firms	Establishments	Business Units	Firms	Segments	Firms
1989	192,908	13,227	2,680	1,481	6,063,857	5,093,300	4,978,250	6,203	3,929
					(3.2%)	(0.26%)	(0.03%)	(43%)	(38%)
1990	197,606	13,167	2,787	1,563	6,126,016	5,136,656	5,024,252	6,000	3,754
					(3.2%)	(0.26%)	(0.03%)	(46%)	(42%)
1991	211,603	13,688	3,011	1,762	6,155,181	5,117,974	5,005,347	5,857	3,687
					(3.4%)	(0.27%)	(0.04%)	(51%)	(48%)
1992	214,973	14,269	3,269	1,999	6,275,349	5,172,296	5,051,405	5,879	3,745
					(3.4%)	(0.28%)	(0.04%)	(56%)	(53%)
1993	197,253	12,559	2,801	1,700	6,356,799	5,271,301	5,149,208	5,686	3,665
					(3.1%)	(0.24%)	(0.03%)	(49%)	(46%)
1994	183,626	11,729	2,558	1,549	6,465,057	5,352,915	5,232,956	5,585	3,649
					(2.8%)	(0.22%)	(0.03%)	(46%)	(42%)
1995	177,937	11,006	2,340	1,403	6,566,634	5,439,734	5,322,981	5,470	3,620
					(2.7%)	(0.20%)	(0.03%)	(43%)	(39%)
1996	179,465	10,641	2,121	1,251	6,699,635	5,555,476	5,439,206	5,318	3,528
					(2.7%)	(0.19%)	(0.02%)	(40%)	(35%)
All firm-years	1,555,371	100,286	21,567	12,708	50,708,528	42,139,652	41,203,605	45,998	29,577
					(3.1%)	(0.24%)	(0.03%)	(47%)	(43%)

Table II

Number of Firm-Year Observations by Number of Segments or Business Units

The table reports the number (percentage) of firms in the sample that have one, two, or more segments (or business units). Segments and business units refer to the operations of a firm in a particular industry, according to COMPUSTAT and BITS, respectively. A *segment* is defined as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit." A *business unit* is the aggregation of all of a firm's establishments with a common four-digit SIC code. An establishment is defined in BITS as "a single physical location where business is conducted or where services or industrial operations are performed." The figures reported for COMPUSTAT refer to all those active and inactive ("research") firms that appear in both the company and the segment COMPUSTAT files, and for which market values can be computed. The Common Sample refers to the merger of the BITS and COMPUSTAT, as described in the Appendix.

Number of Segments or Business Units in Firm	Common Sample				COMPUSTAT (Segments)	
	Business Units		Segments			
	No. of Firm-Years	(Percentage)	No. of Firm-Years	(Percentage)	No. of Firm-Years	(Percentage)
1	2,616	(20.6)	8,114	(63.8)	21,115	(71.4)
2	1,494	(11.8)	2,025	(15.9)	3,815	(12.9)
3	1,367	(10.8)	1,480	(11.6)	2,605	(8.8)
4	1,084	(8.5)	715	(5.6)	1,301	(4.4)
5	870	(6.8)	236	(1.9)	425	(1.4)
6	694	(5.5)	90	(0.7)	200	(0.7)
7	567	(4.5)	29	(0.2)	71	(0.2)
8	521	(4.1)	3	(0)	12	(0.04)
9	384	(3)	4	(0)	14	(0.05)
10	320	(2.5)	12	(0.1)	19	(0.06)
11–20	1,675	(13)				
21–30	636	(5.1)				
31+	480	(3.2)				
Total	12,708	(100)	12,708	(100)	29,577	(100)

Table II reports the number of firm-year observations in the Common Sample and in COMPUSTAT by number of business units, segments, or both. Of special relevance is the comparison between the business unit and segment breakdowns within the Common Sample, that is, between columns one through four. For instance, as many as 43 percent of the sample firms switch from the undiversified to the diversified group when one moves from a segment breakdown of firms to a business unit breakdown. The table also indicates that the right-censoring imposed by COMPUSTAT on measured diversification is binding for a considerable number of firms: 21 percent of the firms have more than 10 business units.

Table III shows additional descriptive statistics about the Common Sample as compared to COMPUSTAT firms, including average employment, assets, and Tobin's q in each year. The table shows that the firms in the Common Sample are comparable to those in COMPUSTAT: they are somewhat larger when

Table III
Firm Employment, Assets, and Tobin's *q*: Means and Standard Deviations

Tobin's q is computed as the market value of common equity plus total assets minus the book value of common equity, divided by total assets. Data on both assets and q come from COMPUSTAT. Employment data come from both COMPUSTAT and BITS. COMPUSTAT employment figures represent the number of company workers as reported to shareholders, which for some firms is the average number of employees over the year and for others the number of employees at year-end. The figures include all employees of consolidated subsidiaries, both domestic and foreign, all part-time and seasonal employees, full-time equivalent employees, and company officers. BITS employment figures are for U.S.-based employees only and include full-time, part-time, and temporary employees, and salaried personnel. Employment in BITS is measured in the pay period that includes March 12 of every year. COMPUSTAT here refers to all active and inactive ("research") firms included in both the company and the segment COMPUSTAT files, and for which market values can be computed. The Common Sample refers to the merger of BITS and COMPUSTAT, as described in the Appendix. A *segment* is defined as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit." A *business unit* is the aggregation of all of a firm's establishments with a common four-digit SIC code. An establishment is defined in BITS as "a single physical location where business is conducted or where services or industrial operations are performed." The sample sizes on which these descriptive statistics are based are shown in Table I. Standard deviations are in parentheses.

	Common Sample				COMPUSTAT		
	Employment						
Year	BITS	COMPUSTAT	Assets ($Million)	q	Employment	Assets ($Million)	q
1989	8,454	10,157	1,748	1.7	7,145	1,468	2.0
	(27,324)	(31,045)	(8,590)	(1.8)	(25,551)	(6,955)	(4.5)
1990	8,040	9,594	1,766	1.6	7,347	1,647	1.9
	(26,749)	(29,659)	(8,963)	(1.8)	(25,668)	(7,706)	(13.2)
1991	7,360	8,901	1,672	1.9	7,663	1,792	2.2
	(24,500)	(28,480)	(8,792)	(2.3)	(25,973)	(8,214)	(9.5)
1992	6,630	8,059	1,601	1.9	7,651	1,893	2.2
	(23,212)	(26,778)	(9,233)	(1.7)	(25,741)	(9,004)	(10.2)
1993	7,200	8,560	1,797	1.9	7,914	2,220	2.1
	(24,707)	(27,412)	(10,331)	(1.6)	(26,220)	(10,526)	(5.6)
1994	7,651	9,197	1,928	1.7	8,263	2,498	1.8
	(24,981)	(28,647)	(9,866)	(1.1)	(27,496)	(11,607)	(2.3)
1995	8,367	10,049	2,187	1.9	8,735	2,878	2.2
	(26,349)	(30,929)	(11,308)	(1.5)	(28,358)	(13,457)	(5.9)
1996	9,492	11,896	2,668	1.9	9,524	3,378	2.3
	(28,446)	(34,848)	(13,170)	(1.4)	(30,106)	(15,725)	(11.5)
All firm-years	7,791	9,412	1,884	1.8	8,011	2,206	2.1
	(25,638)	(29,532)	(9,999)	(1.7)	(26,902)	(10,724)	(8.6)
Of Which:							
Multisegment	12,910	15,855	3,930	1.5	15,972	4,857	1.5
	(35,920)	(39,745)	(15,807)	(1.0)	(39,637)	(15,768)	(1.1)
Single-segment	4,893	5,765	725	2.0	4,721	1,143	2.3
	(16,609)	(20,854)	(3,379)	(2.0)	(18,346)	(7,584)	(10.1)
Multibusiness	9,726	11,706	2,342	1.6			
	(28,449)	(32,731)	(11,169)	(1.0)			
Single-business	326	563	119	2.6			
	(842)	(2,305)	(718)	(3.1)			
Single-segment but multibusiness	6,775	7,917	983	1.7			
	(19,424)	(24,417)	(3,963)	(1.1)			

size is measured by employment (9,412 vs. 8,011 employees), but smaller when size is measured by assets (1,884 vs. 2,206 millions of dollars). Table III also reveals some discrepancy between BITS and COMPUSTAT in the employment figures for the Common Sample. There are two main reasons for this discrepancy. First, BITS and COMPUSTAT use different measures of employment. These measures are described in detail in the Appendix. Second, if firms have non-U.S. operations, they will be consolidated in the balance sheet and their employees included in the employment data item in COMPUSTAT. In contrast, BITS includes only U.S. establishments and their employees. This second reason may raise a broader concern about the coverage of certain firms in the sample by BITS data. The Appendix also provides a detailed analysis of the extent to which a firm's operations are covered. The analysis shows that coverage ratios are uncorrelated with the key variables in this study, particularly with excess values. For this reason, the main analysis uses all firms in the sample, regardless of the percentage of their operations that are in the United States. In Section II.C, I examine the robustness of the main results to the exclusion of firms with non-U.S. operations from the sample.

B. Excess Value Measures

The two most important variables for this study are excess value and diversification. Following prior literature, diversification is measured by a multisegment or a multibusiness dummy, depending on the data. Excess value is measured in two different ways. First, following Lang and Stulz (1994), Servaes (1996), and Rajan, Servaes, and Zingales (2000), it is computed as the difference between a firm's Tobin's q and its imputed q. Second, following Schoar's (2002) adaptation of Berger and Ofek's (1995) measure, excess value is computed as the natural logarithm of the ratio of Tobin's q to the imputed q. This second measure is used to facilitate the comparison with these and other studies that have used Berger and Ofek's approach.

Tobin's q is proxied by the ratio of the market value of common equity plus the book value of preferred stock and debt to total assets. Because Tobin's q only uses firm-level data, it is the same for any given firm, regardless of whether excess values are estimated using BITS or using segment data. On the other hand, *imputed* qs are contingent on whether diversification is measured using BITS or using segment data. In segment data, the imputed q is the asset-weighted average of the hypothetical qs of the firm's segments, where a segment's hypothetical q is the average of the single-segment firms in the industry in the year examined. On BITS data, imputed q is the employment-weighted average of the hypothetical qs of the firm's business units. A business unit's hypothetical q is the average of the single-business firms in the industry.

There are two key differences between the imputed qs based on BITS and those based on segment data: the industry averages and the weights used. Industry average qs are analyzed in detail in the following subsection. The weights used as measures of business unit or segment size for the main analysis are employment in BITS, and assets in COMPUSTAT. Employees are used as weights for business units because it is the only size measure available in BITS. Assets are used as the primary weights for segments to follow the precedent in earlier studies. I also use employment as weights for segments, to test directly whether employment-weighted imputed qs are lower than the asset-weighted ones or not. Section II.B examines the sensitivity of the main results to the use of different weights.

C. Industry Tobin's q Measures

Industry averages on either data source are computed at the most precise SIC level for which there is a minimum of three pure-play firms. The set of pure-play firms in an industry is, however, different in the two databases. This makes the industry average qs different across databases as well. The trade-off between the two sets of averages is the following. For four-digit industries with a large enough number of firms, BITS industry averages are necessarily more accurate than segment averages, because they are based only on the true pure-plays and not on segments that usually lump several different businesses together. However, because BITS is much more disaggregated than COMPUSTAT, there are less pure-play firms available to compute industry averages: Table II shows that single-business firms are only 21 percent of the Common Sample, whereas single-segment firms are 64 percent. As a result, the proportion of industry averages that are computed at broader SIC levels is higher in BITS.

Table IV provides descriptive statistics about the two sets of industry q measures. Panel A reports the number of pure-play firms in each industry that are used to compute average qs. Panel B shows how many of the BITS and COMPUSTAT industry qs are based on four-, three-, two-, or one-digit SIC measures. Also reported are two features of the sample that contribute to the proportion of broadly defined industry averages being higher in BITS than in segment data. One, BITS contains a much higher diversity of SIC codes than COMPUSTAT: The Common Sample covers 889 different four-digit SIC codes according to BITS, but only 655 according to COMPUSTAT (see table footnote). This can be expected, given that there are also many more business units than segments in the sample. Two, there is a higher proportion of SIC codes defined at the three-, two-, and one-digit levels among business units than there is among segments.

Panel C shows how the industry qs based on BITS differ from those based on COMPUSTAT. BITS-based industry qs are on average higher than the COMPUSTAT-based ones (2.48 vs. 1.90). This is consistent with the evidence in Table III that the q of single business firms in the Common Sample (2.6) is higher than the q of single-segment firms in the same sample (2.0). The variance across industry qs and over time is also higher in BITS than in COMPUSTAT.

The statistics reported in Table IV confirm that the two sets of industry qs are different. This raises the question of which measures are more accurate. Table V reports the results of two analyses that throw some light onto this question. I first examine whether and how closely the SIC codes reported in COMPUSTAT match those in BITS. This analysis is carried out only on the subsample of firms that are classified as single-segment in COMPUSTAT, for two reasons:

Table IV

Industry Tobin's qs: Descriptive Statistics

Industry Tobin's q on BITS (COMPUSTAT) is computed as the average q of all single-business (single-segment) firms in the industry in any given year. Segments and business units refer to the operations of a firm in a particular industry according to COMPUSTAT and BITS, respectively, and are defined in the footnotes to Tables I and II. Averages are computed at the most precise SIC level for which there is a minimum of three single-business (single-segment) firms in the industry. Tobin's q is computed as the market value of common equity plus total assets minus the book value of common equity, divided by total assets. The sample is the BITS–COMPUSTAT Common Sample. The observations to which the percentages below refer are all the business unit-years or segment-years in the Common Sample ($N = 100{,}286$ for BITS-based industry qs; $N = 21{,}567$ for COMPUSTAT-based industry qs). Equivalently, the sample size is the number of different SIC codes in the sample weighted by the frequency with which each industry average is used in the computation of excess values. The Common Sample includes 889 different four-digit SIC codes in BITS and 655 in COMPUSTAT.

Panel A. Number of Pure-Play Firms in Industry Used to Compute Industry Average q

	3–5	6–10	11–20	21+	Total
BITS industry qs	30	16	14	40	100
COMPUSTAT industry qs	20	20	20	40	100

Panel B. Percentage of Industry qs, Business Units, and Segments, by SIC Level

	4-digit	3-digit	2-digit	1-digit	Total
BITS industry qs	13	13	40	34	100
COMPUSTAT industry qs	40	21	31	8	100
Business units defined at each SIC level in BITS	84	5	10	1	100
Segments defined at each SIC level in COMPUSTAT	98	2	0.3	0.2	100

Panel C. Summary Statistics for Industry qs

	Mean	*SD*	Within-Industry *SD*/Mean
BITS industry qs	2.48	1.13	0.25
COMPUSTAT industry qs	1.90	0.95	0.20

one, these are the only firms used to compute industry qs in COMPUSTAT, and they in turn include all the single-business firms used to compute industry qs in BITS. Two, these are the only firms for which BITS and COMPUSTAT can be matched at the segment level in an objective way.

Panel A shows the results of this exercise broken down into single-business firms and multibusiness (but still single-segment) firms. The table shows that approximately half (51 percent) of the single-segment, single-business firms have the same four-digit SIC code in both BITS and COMPUSTAT. One fourth have SIC codes that match at higher levels of aggregation: three-digit (9 percent), two-digit (8 percent), or one-digit (8 percent). The remaining fourth of this subsample has SIC codes that do not match even at the broadest (sector) level of industry definition.

The analysis of the single-segment-but-multibusiness firms shows that 84 percent of these firms choose their single segment's SIC code from among the set of four-digit codes in which they are present according to BITS. Only 3

Table V

Quality of the BITS and COMPUSTAT Industry *q*s

Industry Tobin's *q* on BITS (COMPUSTAT) is computed as the average *q* of all single-business (single-segment) firms in the industry in any given year. A *segment* is defined as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit." A *business unit* is the aggregation of all of a firm's establishments with a common four-digit SIC code. An establishment is defined in BITS as "a single physical location where business is conducted or where services or industrial operations are performed." Averages are computed at the most precise SIC level for which there is a minimum of three single-business (single-segment) firms in the industry. Tobin's *q* is computed as the market value of common equity plus total assets minus the book value of common equity, divided by total assets. The sample in Panel A is the subsample of 8,114 single-segment firms in the BITS–COMPUSTAT Common Sample. These include 2,616 single-business firms and 5,498 multibusiness firms. The sample in Panel B is the subsample of "ideal" industries, that is, all the four-digit industry codes for which there are three or more single-business firms. These firms are pure-plays according to both BITS and COMPUSTAT. The number of observations is 12,914 business units (or single-business firms) from the "ideal" industries. The asterisk indicates a 1 percent significant difference from the "ideal" *q*.

Panel A. Maximum Level of Precision at which BITS and COMPUSTAT SIC Codes Match

	4-digit	3-digit	2-digit	1-digit	None	Total
Single-segment, single-business firms	51	9	8	8	24	100
Single-segment but multibusiness firms:						
Segment SIC Code Matches Any Business Unit's SIC Code	84	6	4	3	3	100
Segment SIC Code Matches Largest Business Unit's SIC Code	14	1	2	2	81	100

Panel B. Comparison of Ideal *q*s and *q*s Computed at Different Aggregation Levels in BITS/COMPUSTAT

	4-digit	3-digit	2-digit	1-digit
BITS	*Ideal: 2.56*	2.55	2.61*	2.57
COMPUSTAT	2.04*	2.04*	2.05*	1.94*

percent choose a segment SIC code that does not match one of their business units' SIC codes at the three, two, or one-digit level. However, only 14 percent of firms choose to report the four-digit SIC code of their largest business unit. If business units are aggregated to the three-, two-, or one-digit level, only an additional 5 percent of firms choose to report as a segment the SIC code of their largest business. That is, in 81 percent of cases, firms choose a segment SIC code that does not match that of its largest business, even when businesses are defined at the broadest possible level.

Panel B of Table V shows how the industry *q*s are affected by the inclusion of unrelated activities into COMPUSTAT segments. The relative accuracy of the BITS and COMPUSTAT industry *q*s is compared by focusing on the subsample of four-digit industry codes for which there are three or more pure-play firms according to BITS. This subsample contains the "ideal" industry *q*s, in the sense that they are the only ones based on a reasonable number of truly pure-play firms. The results show that industry averages computed at higher levels of aggregation on BITS data (particularly at the three- or one-digit level) are not significantly different from the ideal averages. In contrast, the COMPUSTAT averages for the same industries differ significantly from the ideal ones.

II. Results

A. Do the Results on Segment Data Obtain on Business Unit Data?

To ensure that my sample is comparable to those in prior studies, I first verify whether the finding of a discount on segment data also holds for the Common Sample. Following Lang and Stulz (1994), Servaes (1996), and Rajan, Servaes, and Zingales (2000), I estimate the average excess value of multisegment firms relative to single-segment firms as the mean difference between the two.

Table VI reports the estimates obtained. The first column shows that, consistent with prior studies, the multisegment firms in my sample trade at a statistically significant discount relative to their single-segment counterparts.

Table VI
Diversified Firms' Excess Value in the BITS–COMPUSTAT Common Sample

Diversified firms' excess value is computed as the mean difference in individual excess values between diversified (multisegment or multibusiness) firms and nondiversified (single-segment or single-business) firms. A *segment* is defined as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit." A *business unit* is the aggregation of all of a firm's establishments with a common four-digit SIC code. An establishment is defined in BITS as "a single physical location where business is conducted or where services or industrial operations are performed." Individual firms' excess values are computed as the difference between the firm's Tobin's q and its imputed q. Tobin's q is computed as the market value of common equity plus total assets minus the book value of common equity, divided by total assets. A firm's imputed q is the size-weighted average of the hypothetical qs of its segments (business units). Segment size is measured by assets; business unit size is measured by employment. A segment's (business unit's) hypothetical q is the average q of all single-segment (single-business) firms in its industry in any given year. Industry averages are computed at the four-digit SIC code level whenever possible. The sample is the BITS–COMPUSTAT Common Sample.

		Segment Data (COMPUSTAT)			Business Unit Data (BITS)		
Year	No. of Firms	Mean	(t-Stat)	% Firms Diversified	Mean	(t-Stat)	% Firms Diversified
1989	1,481	−0.29	(−4.09)	41	0.24	(2.11)	82
1990	1,563	−0.11	(−1.75)	40	0.40	(4.30)	80
1991	1,762	−0.21	(−2.63)	37	0.11	(0.77)	78
1992	1,999	−0.19	(−2.86)	34	0.29	(3.57)	74
1993	1,700	−0.18	(−2.48)	34	0.28	(2.90)	76
1994	1,549	−0.09	(−1.56)	15	0.43	(5.36)	80
1995	1,403	−0.22	(−2.59)	35	0.20	(1.63)	82
1996	1,251	−0.15	(−1.61)	37	0.35	(2.52)	86
All	12,708	−0.18	(−6.92)	36	0.28	(7.31)	79

The mean discount ranges between −0.09 and −0.29 for different years between 1989 and 1996, and averages −0.18 over the whole period. These figures are smaller than those reported by Lang and Stulz (1994) and Servaes (1996) for the 1980s and the 1960s, respectively. The results are consistent, however, with the general downward trend in the size of the diversification discount reported by Lang and Stulz (1994) for the late 1980s.

Column three of Table VI reports the corresponding estimates obtained on business unit data. As shown, when excess values are estimated in the same sample using BITS data, diversified firms trade at an average *premium* relative to single-business firms. The premium ranges between 0.11 and 0.43, with a 1989 to 1996 period average of 0.28. In other words, the finding of a "diversification discount" gets completely reversed when a more consistent and objective definition of diversified firms' constituent units is used. This suggests that one should be wary of inferences about the effect of diversification on firm value based only on segment data, since the sign of the effect is contingent on the data source used. Before reaching any further conclusion, however, a number of robustness checks need to be performed.

B. Robustness of the Results to Variations in the Computation of Excess Values

In this subsection, I examine the sensitivity of the results to variations in the measurement of excess values. The results of the robustness checks are reported in Table VII. To facilitate the comparison, the first row of Table VII reproduces the last row of Table VI (the main pooled results), and only the pooled estimates are reported for each variation. Results for individual years are similar and are available from the author upon request.

The first robustness check analyzes the sensitivity of the main results to the measures of industry q. I exclude from the sample those firms with at least one business unit for which the average industry q in BITS could not be computed at any SIC level other than the one-digit (sector) level. Because this affects the composition of the sample, excess values are also recomputed on segment data for comparison. Row two of Table VII shows a significant discount of −0.25 on segment data, and a smaller, but large and significant, premium of 0.18 on BITS data.

The next two robustness checks deal with the issue of the weights attached to segments and business units in the computation of imputed qs. In the main analysis reported in Table VI and in row one of Table VII, assets are used as weights for segments (following earlier studies), but employment figures are used as weights for business units (because it is the only available measure of business unit size). If employment-weighted imputed qs are systematically higher than asset-weighted imputed qs, this will explain at least part of the difference between the discount on segment data and the premium on BITS. To test for this possibility, excess values are re-estimated on segment data using segment employment as weights. Row three of Table VII shows, however, that the discount at which the multisegment firms in the sample trade relative to single-segment firms is even larger if employment is used (−0.22 vs. −0.18).

Table VII

Results of Robustness Checks

Diversified firms' excess value is computed as the mean difference in individual excess values between diversified (multisegment or multibusiness) firms and nondiversified (single-segment or single-business) firms, pooling all firm-year observations in the sample. A *segment* is defined as "a component of an enterprise engaged in providing a product or service or a group of related products and services primarily to unaffiliated customers (i.e., customers outside the enterprise) for a profit." A *business unit* is the aggregation of all of a firm's establishments with a common four-digit SIC code. An establishment is defined in BITS as "a single physical location where business is conducted or where services or industrial operations are performed." Individual firms' excess values are computed as the difference between the firm's Tobin's q and its imputed q. Tobin's q is computed as the market value of common equity plus total assets minus the book value of common equity, divided by total assets. A firm's imputed q is the size-weighted average of the hypothetical qs of its segments (business units). Segment size is measured by assets and business unit size is measured by employment, unless otherwise indicated. A segment's (business unit's) hypothetical q is the average q of all single-segment (single-business) firms in its industry in any given year. Industry averages are computed at the four-digit SIC code level whenever possible. The sample is the BITS–COMPUSTAT Common Sample. The number of observations is 12,708 unless otherwise indicated. The t-statistics are in parentheses.

	Segment Data			Business Unit data		
Robustness Check	Mean	(t-Stat)	% Firms Diversified	Mean	(t-Stat)	% Firms Diversified
Excess value estimates in Table IV	−0.18	(−6.92)	36	0.28	(7.31)	79
Variations in the computation of excess values						
Excluding firms with at least one business unit for which the industry q can only be computed at the one-digit level ($N = 5{,}705$)	−0.25	(−4.37)	36	0.18	(3.68)	79
Using segment employment as weights	−0.22	(−4.66)	36			
Using beginning-of-period employment as weights				0.13	(2.01)	79
Subsamples						
Firms with 100% operations in the United States ($N = 8{,}321$)	−0.21	(−6.48)	31	0.23	(5.11)	74
Pure manufacturing firms ($N = 2{,}332$)	−0.36	(−4.28)	19	0.03	(0.34)	49
Variations in the construction of business units						
Applying 10% materiality condition to business units				0.50	(18.24)	61
Including all vertically related activities in the same unit				0.31	(8.13)	75

In this sense, the premium found in BITS using employment weights can be considered a conservative estimate.

The use of employment as a measure of business unit size may also raise a concern because employment is a variable input factor and moves pro-cyclically with the business cycle. This produces a positive bias for firms that can adjust their level of employment faster. If these firms tend to be diversified, employment weights will lead to a lower estimate of the diversification discount. To address this potential concern, I perform an additional robustness check using beginning-of-period employment weights for the calculation of excess values based on BITS. Row four of Table VII shows that the resulting average excess value is 0.13, which is lower than the main result, but still positive and significant.

C. Robustness of the Results to Variations in the Sample

This subsection examines whether the main results hold if the sample is restricted in two different ways. First, I examine the subsample of firms that have 100 percent of their operations in the United States. As explained in the data section, COMPUSTAT data represent a firm's total and consolidated operations, including non-U.S. operations when firms have them. In contrast, BITS data are based on the firm's U.S. operations only, since the Census Bureau and the other U.S. institutions at the source of this database can only survey domestic establishments. This could potentially create a large discrepancy in the coverage of segments between the two data sets by understating the size of these segments that expand internationally. If diversified firms are more likely to globalize their better business units, the imputed qs of these units in the excess value measures would be understated and thus create bias towards finding a premium.

The results reported in row five of Table VII reveal that this is not the case. The average effect of diversification on the value of firms that are 100 percent U.S.-based is still positive when estimated on BITS data (0.23), and still negative when estimated on segment data (−0.21). This finding is consistent with the evidence provided in the Appendix that although firms with non-U.S. operations have indeed lower coverage ratios than those that are 100 percent U.S.-based, coverage ratios are uncorrelated with excess values.

I next analyze a subsample of pure manufacturing firms, that is, firms with no segments or business units out of the 2000 to 3999 SIC code range. Using a similar sample, Schoar (2002) compares the average diversification discount obtained from COMPUSTAT segment data to that based on LRD data (aggregating establishments within a firm by SIC code in the same way the business units in this paper are constructed). Schoar finds the excess value of diversified firms to be smaller when estimated on LRD data, but still negative and significant, which is particularly puzzling given that she finds a productivity premium for the same firms.

The results of my own analysis of a subsample of pure manufacturing firms are reported in row six of Table VII. Somewhat consistently with Schoar's results, I find that pure manufacturing firms are the only subset of my sample where the diversification premium is relatively small and statistically insignificant. My use of BITS as opposed to LRD, however, reveals that when establishments and firms from all sectors of the economy are included, the diversification discount disappears. The analysis of the same subsample on segment data also yields a much higher discount than that found in the main analysis (−0.36). This suggests that the net benefits of diversification are lower in manufacturing than in other economic sectors.

D. Robustness of the Diversification Premium to Different Measures of Excess Value and Diversification, and to the Inclusion of Control Variables

As an additional robustness check on my results, I estimate pooled univariate and multivariate regressions of the two different excess value measures described before (difference or ratio) for six different diversification measures.

The use of a multisegment or a multibusiness dummy as the primary measure of diversification in this study follows prior research about the diversification discount. Nonetheless, because this measure entails a rather simplistic view of diversification, several other measures are used to validate the results. These other measures include the number of business units in the firm, and four different continuous measures that take higher values with higher degrees of diversification: one minus the Herfindahl index, and the three measures of entropy—total, related, and unrelated.[10] These continuous measures are described in detail in Jacquemin and Berry (1979) and are standard in the strategy and economics literature on diversification. Following Berger and Ofek (1995), three control variables are also included in the multivariate regressions of excess value on diversification: firm size (measured as the natural logarithm of assets), EBIT-to-sales, and capital expenditures-to-sales.

The results of these regressions are reported in Table VIII. The table shows that the size of the diversification premium varies depending on the measures and specification used, but the premium is always positive and significant.

III. Explaining the Diversification Premium

A. The Measurement of Diversification in BITS and COMPUSTAT

The differences in results obtained from BITS and COMPUSTAT can be traced to the measurement of diversification in the two databases. Measured diversification affects measured discounts or premia in two ways: directly, because some firms that are single-segment in COMPUSTAT are actually diversified according to BITS; and indirectly, though the calculation of industry average qs and firm imputed qs. This raises the question of which measures of diversification are preferable from a research standpoint. On the one hand, issues of strategic accounting and inconsistency in segment definitions across firms and over time make segment data less desirable than BITS. On the other hand, it may be argued that if managers have better information than

[10] If P_i is the proportion of a firm's assets in industry i, the Herfindahl index of diversification is $\mathrm{H} = \sum_i P_i^2$, and the total entropy measure is $\mathrm{E_T} = \sum_i P_i^2 \ln(1/P_i)$. Both measures are computed at the four-digit SIC level. Unrelated entropy, $\mathrm{E_U}$, is defined like $\mathrm{E_T}$ but computed at the two-digit SIC level. Related entropy is defined as $\mathrm{E_R} = \mathrm{E_T} - \mathrm{E_U}$.

Table VIII

Robustness of the Diversification Premium to Different Measures of Excess Value and Diversification, and to the Inclusion of Control Variables

This table reports the coefficients of the diversification variable from pooled OLS regressions of excess value on diversification. Multivariate regressions include three control variables: Log of assets; EBIT-to-sales; and capital expenditures-to sales. Individual firms' excess values are computed as the difference or the natural logarithm of the ratio of a firm's Tobin's q to its imputed q. Tobin's q is computed as the market value of common equity plus total assets minus the book value of common equity, divided by total assets. A firm's imputed q is the size-weighted average of the hypothetical qs of its segments (business units). Segment size is measured by assets; business unit size is measured by employment. A segment's (business unit's) hypothetical q is the average q of all single-segment (single-business) firms in its industry in any given year. Industry averages are computed at the four-digit-SIC code level whenever possible. The Herfindahl index of diversification is $\mathrm{H} = \sum_i P_i^2$, and the total entropy measure is $\mathrm{E_T} = \sum_i P_i^2 \ln(1/P_i)$, where P_i is the proportion of a firm's assets in industry i. Both measures are computed at the four-digit SIC level. Unrelated entropy, $\mathrm{E_U}$, is defined like $\mathrm{E_T}$ but computed at the two-digit SIC level. Related entropy is defined as $\mathrm{E_R} = \mathrm{E_T} - \mathrm{E_U}$. The sample is the BITS–COMPUSTAT Common Sample. The number of observations is 12,708 firm-year observations. The t-statistics are in parentheses.

	Measures of Diversification					
	Dummy	No. of Bus. Units	One Minus Herfindahl	Total Entropy	Related Entropy	Unrelated Entropy
Excess value measured as difference						
Univariate regressions	0.28	0.05	0.82	0.99	1.17	0.47
	(7.3)	(11.6)	(16.3)	(11.9)	(10.9)	(4.3)
Multivariate regressions	0.25	0.05	0.90	0.96	1.12	0.41
	(5.7)	(10.7)	(15.9)	(10.8)	(9.9)	(3.8)
Excess value measured as ratio						
Univariate regressions	0.47	0.06	0.97	1.23	1.35	0.69
	(29.8)	(35.6)	(53.1)	(37.6)	(31.2)	(15.0)
Multivariate regressions	0.38	0.02	0.96	1.08	1.13	0.58
	(21.0)	(26.5)	(45.8)	(31.1)	(24.6)	(13.1)

the econometrician about the strategic extent of diversification, COMPUSTAT-based measures would be more meaningful than BITS. Scharfstein (1998), for example, discusses the limitations of the standard approach of classifying businesses as unrelated if they operate in different SIC codes. Two businesses in different SIC codes can in fact produce complementary products or services, or be vertically related.

Although the criteria used by managers to aggregate activities into segments are to some extent unobservable, certain aspects of the institutional definition of a segment can be replicated in the construction of BITS business units to make these units more comparable to COMPUSTAT segments. One of these aspects is the condition that a segment represent at least 10 percent of the firm's assets, sales, or profits for it to be reported as such. For instance, if 95 percent of a firm's activity is in industry A and the remaining 5 percent is in industry B, it may seem more reasonable to classify the firm as nondiversified (as would be done in COMPUSTAT) than as diversified (as would be done with the business units I have constructed using BITS data). Accordingly, I first reconstruct the business units in the Common Sample by imposing on them a 10 percent materiality condition similar to COMPUSTAT's (in this case applied to employment), and I estimate the effect of diversification as before. To avoid distorting the weights, I substitute the within-firm sum of the new business units' employment for the firm's total employment figure.

The results of this analysis are reported in row seven of Table VII. I find a premium that is larger than the one stemming from the original definition of business units (0.50 vs. 0.28). In fact, it is the largest of all premia found in my robustness checks. This finding indicates that the materiality condition, however reasonable it may be, is not responsible for the observed differences in results between BITS and COMPUSTAT.

The requirement for firms to group all vertically related activities into a common segment raises a similar concern. For instance, if A is an input for B, consider the case of a firm that has 50 percent of its assets in industry A and 50 percent in industry B and transfers most of its A-output internally (as opposed to selling it to outside customers). The firm would be classified as diversified according to BITS data, but as nondiversified in COMPUSTAT. A fundamental difference between the two data sources is that BITS treats vertical integration as a form of corporate diversification, whereas COMPUSTAT does not. Put differently, BITS business unit data lead the researcher to estimate the value effect of diversification in operations, whereas COMPUSTAT segment data lead the researcher to estimate the effect of diversification in markets. Clearly, these are two different concepts, and they need not be correlated. A vertically integrated firm may be highly diversified in its operations, but very narrowly focused on one market. On the other hand, a very outsourcing-oriented firm may be highly diversified in its markets, but very narrowly focused on one type of operation (e.g., product design).

Because there is no obvious reason to prefer investigating one of the two types of diversification or the other, as an additional robustness check I reconstruct BITS business units so that all potentially vertically related activities within the firm are included within a common business unit. Because BITS provides no information about internal transfers, I cannot determine the true extent of firms' vertical integration. Instead, I use input–output data to construct an inter-industry vertical relatedness table. Following Matsusaka (1993) and Schoar (2002), each pair of industries is considered vertically related if the industries receive 5 percent or more of their inputs or supply 5 percent or more of their output to each other. In my re-constructed business units, every pair or group of activities in a firm that are related according to this definition are considered a single business unit. The unit is assigned the SIC code of the activity that has the largest number of employees among those included in it.[11] The last row of Table VII shows that the estimate that results from this

[11] Results are similar if the business unit that groups several vertically related units is assigned the SIC code of the most downstream business.

redefinition of business units is a 0.31 premium, higher again than the baseline of 0.28. This higher premium is consistent with Fan and Lang's (2002) finding that the COMPUSTAT-based discount to vertical integration is higher than the discount observed for other forms of diversification. Nevertheless, my results show that the finding of a premium on BITS data is robust to the exclusion of vertical integration as a form of diversification.

B. Remaining Explanations: Relatedness and Strategic Accounting

The results of the various sensitivity analyses discussed above eliminate several candidate explanations for the results: inaccurate industry *q*s in BITS, use of business unit employment as weights, incomplete coverage of firms with non-U.S. operations, measurement of diversification as a dummy variable, lack of control variables in the regression of excess value on diversification, requirement of a 10 percent materiality condition on COMPUSTAT segments to be considered as such, and treatment of vertical integration as a form of diversification.

There are two plausible explanations for this result that cannot be ruled out. These can be referred to as the "relatedness explanation" and the "strategic accounting explanation." The relatedness explanation is based on the notion that the two databases are measuring different concepts of diversification. By definition, a segment is or may be an aggregation of related activities. As a result, measures of diversification based on COMPUSTAT data capture purely conglomerate, or unrelated, diversification. In contrast, measures of diversification based on BITS capture a broader concept of diversification. This broader concept includes both unrelated and related diversification, and even vertical integration.

According to the relatedness explanation, the findings in this paper would indicate that there is a "conglomerate discount," but at the same time there is a premium to related diversification. Because related diversification is relatively more prevalent than purely unrelated diversification, when all diversification types are pooled together as they are in BITS, the resulting effect on firm value is a premium. The overall effect can only be estimated on BITS data but not on COMPUSTAT data, where relatedness is largely unobserved. This explanation is consistent with the cumulative evidence for a nonmonotonic relationship between diversification and accounting profitability (Rumelt (1974), Barney (1996)). It is also partially consistent with Berger and Ofek's (1995) finding that the diversification discount is lower for relatively related diversifiers.[12] Under the relatedness explanation, the different measurements of diversification and

[12] Fan and Lang (2002) use input–output data to examine how relatedness among segments affects the diversification discount. They find that the discount is not smaller for firms with greater complementarity among their segments, and is in fact larger for firms whose segments are vertically related. Note, however, that all vertically related activities of a firm are by definition subsumed in the segment of the most downstream activity. The definition also allows firms to group complementary activities into the same segments. Hence, input–output data are unlikely to capture the true relatedness of a firm's activities, which is mostly unobservable in COMPUSTAT.

its effect on corporate value that result from each data source would be better interpreted as complements than as substitutes.

The strategic accounting explanation suggests a less neutral view of the two data sources. Game-theoretic models of a firm's disclosure choices in the presence of a competitor suggest that high-performing firms are less prone than low performers to disclose financial information, particularly at the segment level (Darrough and Stoughton (1990), Feltham, Gigler, and Hughes (1992)). Also, diversified firms aggregate their activities into segments so as to avoid disclosing information to potential competitors about which of their operations are most lucrative (Hayes and Lundholm (1996)). According to the strategic accounting explanation, the discount would arise in segment data because diversified firms aggregate segment information in ways that may make them appear to be worse performers than single-segment firms in the same industries. This explanation is consistent with the results of empirical studies, such as Harris (1998), Piotroski (2000), and Berger and Hann (2002, 2003), which confirm that managerial segment reporting conforms to the predictions of strategic accounting theory.

The analysis in Table V of how closely the SIC codes of single-segment firms in COMPUSTAT match those in BITS throws some light onto this issue (as well as on the question previously discussed of which industry averages are more accurate). To recall, Table V shows that a significant number of firms choose a segment SIC code that does not match the SIC code of its largest business, even at the broadest possible level of industry definition.[13] This finding suggests that the relatedness and strategic accounting stories are each responsible for part of the observed differences in results. The difference cannot be apportioned between the two explanations, for at least three reasons. First, the matching between BITS and COMPUSTAT SIC codes can only be done in an objective way for single-segment firms, but not for multisegment firms. For instance, if a firm has 133 business units in BITS but only 10 segments in COMPUSTAT, it is unclear which business units are being lumped into what segment. Second, as noted before, managerial motives and information for reporting segment data are partly unobservable to the econometrician. Third, relatedness and strategic accounting are not mutually exclusive criteria for segment reporting even within the same firm. Therefore, it seems plausible that both explanations account for the observed differences in results between BITS and COMPUSTAT.

IV. Conclusion

This paper explores whether the finding that diversified firms in the United States trade at a discount relative to specialized firms may be an artifact of

[13] I also examined whether firms choose to report SIC codes of industries in which the average q is higher or lower than the q of their industries in BITS. No clear selection criterion is observed, however. Firms are almost evenly split between higher-q and lower-q industries in COMPUSTAT relative to BITS.

the data used to construct it. I use a new establishment-level database (BITS), which covers the whole U.S. economy between 1989 and 1996, to construct alternative and arguably better measures of diversification. These data allow me to determine the breakdown of firms' activities by industry in a consistent way across firms. In contrast, the segment data used in prior studies of the diversification discount typically group into each segment different activities, and different combinations of activities for different firms. More importantly, my data are not subject to the strategic accounting biases characteristic of segment reporting prior to 1997. The use of this new data source in a sample where segment data yield a discount reveals that diversified firms actually trade at a large and statistically significant premium relative to specialized firms in the same industries. The premium is robust to a number of variations in the method, sample, and measures of excess value and diversification used.

The two main candidate explanations for this result are that: (a) COMPUSTAT yields a conglomerate discount that is different but consistent with the premium found in BITS for related diversification and (b) the discount found in COMPUSTAT is the result of strategic accounting practices in managerial segment reporting. Both explanations are plausible, and neither can be ruled out with the available data. Nevertheless, my finding that the use of a more objective and fine-grained source of data reverses the fundamental result about the relative value of diversified firms calls into question much of the received wisdom about the diversification discount. It also calls into question the adequacy of segment data for other research topics in corporate finance, strategy, and accounting.

Appendix

A. Construction and Coverage of the Sample

The BITS–COMPUSTAT *Common Sample* has been constructed by merging both databases to the extent to which this is feasible. Each establishment in BITS is identified yearly by a census file number (CFN) and a permanent number. The permanent number remains unchanged even if the CFN changes as a result of structural, legal, or ownership changes in the business. The legal entities owning the establishments are identified by a federal employer identification number (EIN). Firms, which are "the largest aggregation of business legal entities under common ownership and control," are identified uniquely by a so-called *alpha number*.

EIN is the only common data field that may be used to match firms using a computer program, but doing so involves several issues. First, the COMPUSTAT company files include the firm's (primary) EIN, which is time-invariant for any given firm. Some multiunit firms in BITS comprise more than one EIN. Hence, to retrieve from BITS all of a firm's establishments, one has to retrieve the alpha numbers associated with the firm's primary EIN. Second, BITS includes the establishment's EIN, but only the one it had in 1992. This raises two problems that introduce some additional complication in the data-merging process. One, by matching only on EIN, those firms that did not exist in 1992 are left out of the Common Sample. Two, an establishment's EIN may change over time as a result of changes in its ownership structure or legal form of organization. Thus, a direct matching on the EIN-year would attribute some establishments in some years to the wrong owner. Similar problems arise if one attempts to merge COMPUSTAT and BITS using names. BITS itself contains no names, but establishment names can be retrieved from the Business Master Files of the Standard Statistical Establishment List (SSEL), from which BITS actually derives. However, those files are also only available for 1992 (and earlier census years not in BITS).

In order to deal with these issues, a fairly complicated data merging process has been followed, which is documented in detail in Villalonga (2000). This involves computer-matching on EINs, hand-matching on names, and the use of two auxiliary Census databases—SSEL and the LRD. The process maximizes the number of feasible matches without misallocating establishment-year observations across their owning firms. It also ensures that all U.S.-manufacturing establishments of COMPUSTAT firms are covered in the common sample. Yet the absence of time-varying EINs for nonmanufacturing establishments means that some of these establishments may not have been matched to their owning firms.

Table AI provides a detailed analysis of the extent to which a firm's operations are covered in the sample. Coverage is measured by the ratio of the firm's employment figure in BITS to its employment figure in COMPUSTAT. In Panel A, coverage ratios are reported and broken down by firm type, industry, and industry q. Panel B reports on the relation between coverage ratios and the key variables in this study.

One caveat is in order here. Employment is used to measure coverage because it is the only variable in common between the two data sources and the only basis that can be used to analyze the coverage issue. However, because BITS and COMPUSTAT use different measures of employment, the ratio of the two employment figures is not necessarily indicative of coverage. In COMPUSTAT, employment represents the number of company workers as reported to shareholders. Some firms report the yearly average and others the number of employees at year-end. COMPUSTAT makes no attempt to differentiate between the two bases of reporting. The employment item includes all employees of consolidated subsidiaries, both domestic and foreign, all part-time and seasonal employees, full-time equivalent employees, and company officers. It excludes consultants, contract workers, directors, and employees of unconsolidated subsidiaries.

In BITS, employment figures come from multiple sources: IRS payroll tax returns, Social Security Administration applications for EINs, state unemployment insurance administration records from the Bureau of Labor Statistics, and various Census Bureau Surveys such as the economic censuses conducted every five years, the Annual Survey of Manufacturing, or the Company Organization Survey. Employment includes full-time, part-time, and temporary employees, and salaried personnel. It excludes contractors, volunteers, and proprietors or partners. Employment is measured in the pay period that includes March 12

Table AI

Coverage of Firms in the Common Sample

Coverage ratios (CR) for each firm are computed by dividing the firm's employment figure according to BITS, by its employment figure according to COMPUSTAT. COMPUSTAT employment figures represent the number of company workers as reported to shareholders, which for some firms is the average number of employees over the year and for others the number of employees at year-end. The figures include all employees of consolidated subsidiaries, both domestic and foreign, all part-time and seasonal employees, full-time equivalent employees, and company officers. BITS employment figures are for U.S.-based employees only and include full-time, part-time, and temporary employees, and salaried personnel. Employment in BITS is measured in the pay period that includes March 12 of every year. In this table, "industry" refers to the four-digit SIC code of the firm's largest segment in COMPUSTAT. The sample is the BITS–COMPUSTAT Common Sample. The asterisk denotes statistical significance at the 1 percent level.

Panel A. Coverage Ratios by Firm Type, Industry, and Industry q

	Employment		Coverage Ratio		No. of Observations	
	BITS (#)	COMPUSTAT (#)	Mean (%)	Median (%)	Total	w/CR > 100 (%)
All firm-years	7,791	9,410	123	90	12,708	32
By firm type:						
Diversified (multisegment)	12,910	15,855	111	90	4,594	30
Nondiversified (single-segment)	4,893	5,765	131	91	8,114	33
Diversified (multi-business)	9,726	11,706	132	91	10,092	32
Nondiversified (single-business)	326	563	90	89	2,616	33
Pure manufacturing	903	1,249	94	91	2,332	31
With nonmanufacturing operations	9,339	11,247	130	90	10,376	32
100 of operations in the US	5,664	5,502	131	97	8,321	42
Less than 100 of operations in the US	11,827	16,829	110	74	4,387	13
By industry sector						
Mining and construction	5,099	6,191	120	94	1,301	39
Manufacturing	8,732	11,642	129	89	3,079	28
Manufacturing	6,473	8,026	103	87	4,716	27
Transportation and communication	8,826	8,636	143	98	568	46
Wholesale and retail trade	15,460	16,383	127	96	1,472	39
Lodging and entertainment	4,749	6,234	184	90	10,938	33
Services	4,186	4,235	126	95	479	42
By tndustry q						
1st quintile ($q < 1.30$)	9,197	9,994	114	95	2,542	38
2nd quintile ($1.30 \leq q < 1.52$)	8,521	10,628	104	92	2,547	33
3rd quintile ($1.52 \leq q < 1.82$)	8,660	10,738	118	90	2,534	30
4th quintile ($1.92 \leq q < 2.36$)	7,879	9,705	138	89	2,542	30
5th quintile ($q \geq 2.36$)	4,701	6,000	143	85	2,543	29

Panel B. Correlations between Coverage Ratio and Key Variables

Employment					Excess Value	
BITS	COMPUSTAT	Assets	Firm q	Industry q	BITS	COMPUSTAT
0.079*	−0.026	−0.012	0.013	0.022	0.019	0.005

of every year. It is thus possible for an establishment to have zero employment in BITS, but positive employment implicitly included in COMPUSTAT (e.g., if the establishment's business is seasonal or if it is formed after the March 12 pay period).

Taking this caveat into account, the results in Panel A of Table AI suggest the following. First, the mean and median differences in employment figures between BITS and COMPUSTAT do not seem to be generally due to incomplete coverage of the firm's operations in BITS. Although the median coverage ratio is 90 percent, the mean is 123 percent, which if anything would indicate the opposite. This conflicting pattern is also found in most of the breakdown categories. Second, the differences in coverage ratios between pure manufacturing firms and the rest give no indication of incomplete matching of nonmanufacturing establishments to their owning firms. In fact, the mean coverage is higher for firms with nonmanufacturing operations (130 percent) than it is for pure manufacturers (90 percent), while the median ratios are similar (90 percent and 91 percent, respectively). Third, the coverage of firms with non-U.S. operations is, as expected, smaller than that of purely U.S.-based firms. The mean coverage ratios are 110 percent and 130 percent, respectively, and the median ratios are 74 percent and 97 percent.

In summary, the discrepancy between the employment figures in BITS and COMPUSTAT for the Common Sample seems attributable to the different concepts of employment used by each source, and does not reflect incomplete coverage of the sample firms' operations, except for firms with non-U.S. establishments. Panel B shows that coverage ratios are uncorrelated with the key variables in this study, particularly with excess values.

REFERENCES

Acs, Zoltan, and Catherine Armington, 1998, Longitudinal Enterprise and Establishment Microdata (LEEM) documentation, CES Discussion paper 98-9, Center for Economic Studies, U.S. Bureau of the Census, Washington DC.

Acs, Zoltan, Catherine Armington, and Alicia Robb, 1999, Measuring job flow dynamics in the U.S. economy, CES Discussion paper 99-1, Center for Economic Studies, U.S. Bureau of the Census, Washington DC.

American Institute of Certified Public Accountants (AICPA), 1994, Improving business reporting—A customer focus, Report of the AICPA Special Committee on Financial Reporting, AICPA, New York, NY.

Armington, Catherine, 1998, *Statistics of U.S. Businesses—Microdata and Tables of SBA/Census Data on Establishments by Firm Size* (Office of Advocacy, U.S. Small Business Administration, Washington DC).

Armington, Catherine, and Alicia Robb, 1998, *Mergers and Acquisitions in the U.S.: 1990–1994* (Office of Advocacy, U.S. Small Business Administration, Washington DC).

Armington, Catherine, and Alicia Robb, 1999, *An investigation into the volume and impact of business acquisition activity in the U.S.: 1990–1994* (Office of Advocacy, U.S. Small Business Administration, Washington DC).

Association for Investment Management Research (AIMR), 1993, Financial reporting in the 1990s and beyond: A position paper of the AIMR, prepared by Peter H. Knutson for the Financial Accounting Policy Committee, AIMR, Charlottesville, VA.

Barney, Jay B., 1996, *Gaining and Sustaining Competitive Advantage* (Addison-Wesley, Reading, MA).

Berger, Philip G., and Rebecca Hann, 2002, Segment disclosures, proprietary costs, and the market for corporate control, Working paper, University of Chicago and University of Southern California.

Berger, Philip G., and Rebecca Hann, 2003, The impact of SFAS No. 131 on information and monitoring, *Journal of Accounting Research* 41, 163–223.

Berger, Philip G., and Eli Ofek, 1995, Diversification's effect on firm value, *Journal of Financial Economics* 37, 39–65.

Berger, Philip G., and Eli Ofek, 1996, Bustup takeovers of value-destroying diversified firms, *Journal of Finance* 51, 1175–1200.

Campa, José Manuel, and Simi Kedia, 2002, Explaining the diversification discount, *Journal of Finance* 57, 1731–1762.

Darrough, Masako, and Neal M. Stoughton, 1990, Financial disclosure policy in an entry game, *Journal of Accounting and Economics* 12, 219–243.

Davis, Rachel, and Irene M. Duhaime, 1992, Diversification, vertical integration, and industry analysis: New perspectives and measurement, *Strategic Management Journal* 13, 511–524.

Denis, David J., Diane K. Denis, and Atulya Sarin, 1997, Agency problems, equity ownership and corporate diversification, *Journal of Finance* 52, 135–160.

Fan, Joseph P.H., and Larry H.P. Lang, 2002, The measurement of relatedness: An application to corporate diversification, *Journal of Business* 73, 629–660.

Feltham, Gerald A., Frank B. Gigler, and John S. Hughes, 1992, The effects of line-of-business reporting on competition in oligopoly settings, *Contemporary Accounting Research* 9, 1–23.

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

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

Graham, John R., Michael Lemmon, and Jack Wolf, 2002, Does corporate diversification destroy value? *Journal of Finance* 57, 695–720.

Harris, Mary S., 1998, The association between competition and managers' business segment reporting decisions, *Journal of Accounting Research* 36, 111–128.

Hayes, Rachel M., and Russell Lundholm, 1996, Segment reporting to the capital market in the presence of a competitor, *Journal of Accounting Research* 31, 261–279.

Hyland, David C., 1997, Why firms diversify, An empirical examination, unpublished doctoral dissertation, Ohio State University, Columbus, Ohio.

Jacquemin, Alexis P., and Charles H. Berry, 1979, Entropy measure of diversification and corporate growth, *Journal of Industrial Economics* 27, 359–369.

Lang, Larry H.P., and René M. Stulz, 1994, Tobin's q, corporate diversification, and firm performance, *Journal of Political Economy* 102, 1248–1280.

Lichtenberg, Frank R., 1991, The managerial response to regulation of financial reporting for segments of a business enterprise, *Journal of Regulatory Economics* 3, 241–249.

Maksimovic, Vojislav, and Gordon Phillips, 2002, Do conglomerate firms allocate resources inefficiently? *Journal of Finance* 57, 721–767.

Martin, John D., and Akin Sayrak, 2003, Corporate diversification and shareholder value, *Journal of Corporate Finance* 9, 37–57.

Matsusaka, John G., 1993, Takeover motives during the conglomerate merger wave, *RAND Journal of Economics* 24, 357–379.

Montgomery, Cynthia A., 1994, Corporate diversification, *Journal of Economic Perspectives* 8, 163–178.

Nguyen, Sang V., 1998, The manufacturing plant ownership change database: Its construction and usefulness, CES Discussion paper, Center for Economic Studies, U.S. Bureau of the Census, Washington DC.

Piotroski, Joseph D., 2000, Discretionary segment reporting decisions and the precision of investors beliefs, Working paper, University of Chicago.

Rajan, Raghuram, Henri Servaes, and Luigi Zingales, 2000, The cost of diversity: The diversification discount and inefficient investment, *Journal of Finance* 60, 35–80.

Robb, Alicia, 2000, New data for dynamic analysis: The Business Information Tracking Series (BITS). Office of Advocacy, U.S. Small Business Administration, Washington DC.

Rumelt, Richard P., 1974, *Strategy, Structure, and Economic Performance* (Division of Research, Harvard Business School, Boston, MA).

Scharfstein, David S., 1998, The dark side of internal capital markets II: Evidence from diversified conglomerates, Working paper 6352, National Bureau of Economic Research, Cambridge, MA.

Schoar, Antoinette S., 2002, Effects of corporate diversification on productivity, *Journal of Finance* 57, 2379–2403.

Servaes, Henri, 1996, The value of diversification during the conglomerate merger wave, *Journal of Finance* 51, 1201–1225.

Villalonga, Belén, 1999, Does diversification cause the "diversification discount"? Working paper, Harvard University.

Villalonga, Belén, 2000, Matching BITS to COMPUSTAT: Towards richer data for large sample research within firms, CES Discussion paper, Center for Economic Studies, U.S. Bureau of the Census, Washington DC.