

Large Blocks of Stock: Prevalence, Size, and Measurement

Jennifer Dlugosz
Department of Economics and Harvard Business School
Harvard University

Rüdiger Fahlenbrach
Department of Finance, Fisher College of Business
Ohio State University

Paul Gompers
Harvard Business School
Harvard University and NBER

Andrew Metrick
Department of Finance, Wharton
University of Pennsylvania and NBER

January 2005

We thank Sith Chaisurote, Ashton Hawk, Allison Lowenstein, Jason O'Connor, Colleen Pontious, and Rebecca Yang for excellent research assistance. We thank Ralph Walkling for helpful comments and suggestions. Gompers and Metrick acknowledge support from NSF grant #SES-0136791. Gompers acknowledges support from the Division of Research, Harvard Business School. The data set built in this paper can be downloaded from <http://finance.wharton.upenn.edu/~metrick/data.htm>.

Abstract

Large blocks of stock play an important role in many studies of corporate governance and finance. Despite this important role, there is no standardized data set for these blocks, and the best available data source, *Compact Disclosure*, has many mistakes and biases. In this paper, we document these mistakes and show how to fix them. The mistakes and biases tend to increase with the level of reported blockholdings: in firms where *Compact Disclosure* reports that aggregate blockholdings are greater than 50 percent, these aggregate holdings are incorrect more than half the time and average holdings for these incorrect firms are overstated by almost 30 percentage points. For researchers using uncorrected blockholder data as a dependent variable, these errors will increase the standard error of coefficient estimates but do not appear to cause bias. However, we find that if blockholders are used as an independent variable, economically significant errors-in-variables biases can occur. We demonstrate these biases using a representative analysis of the relationship between firm value and outside blockholders. An online appendix to our paper provides a “clean” data set for our sample firms and time period. For researchers who need to work outside of this sample, we also test the efficacy of alternative (cheaper) fixes to this data problem, and find that truncating or winsorizing the sample can reduce about half of the bias in our representative application.

I. Introduction

Large-block shareholders play an important role in corporate governance. For this reason, the presence of such “blockholders” and the size of their holdings is a common explanatory variable in financial research. In just the last few years, a representative sample of such studies includes analyses of the role of blockholders in executive turnover, executive compensation, firm diversification, discretionary expenses, market liquidity, and corporate performance.¹ Furthermore, blockholder data is a crucial input in the analysis of the relationship between ownership structure and firm value, where seminal works by Demsetz and Lehn (1985) and Morck et al. (1988) gave rise to a vast and growing literature.

Despite the common use of large shareholder data, there is no clean off-the-shelf database to facilitate research. Many of the papers cited above required their authors to gather their own data. This time-consuming task is necessary because of several weaknesses in the available databases. Of course, decentralized data gathering causes duplication of effort and lack of standardization across projects. Also, because of the large time commitment necessary to clean the data for each firm, most researchers have gathered data for a relatively small number of firms. This paper aims to fill this data gap by documenting the problems with the currently available data, proposing a consistent set of solutions to these problems, and making a “clean” database freely available to all researchers.² Furthermore, we demonstrate the superiority of clean (vs. raw) data with a representative study on the relationship between outside blockholders and firm value and discuss some alternatives to this exhaustive cleaning for other samples.

¹ For examples of papers on these listed topics, see Denis et. al (1997), Ryan and Wiggins (2001), Anderson et. al (2000), Ang et. al (2000), Singh and Davidson (2003), Heflin and Shaw (2000), Cremers and Nair (2004), and Shivdasani (1993). Holderness (2003) gives a survey of the blockholder literature.

² The database can be downloaded from <http://finance.wharton.upenn.edu/~metrick/data.htm>

The Securities Exchange Act of 1934 (SEA) lays out the ownership disclosure requirements for public corporations in Regulation 14A and Schedule 14A. Virtually everything we know about blockholders in the United States comes from these disclosure requirements, which are described in detail in Appendix A of this paper. The two main types of data produced by the SEA are for holdings (once per year, reported in the annual proxy statement), and for transactions by corporate insiders and beneficial owners (updated through Forms 3, 4 and 5). While the trading data would appear to provide the most current and comprehensive information, past research has demonstrated that this data is difficult to work with and cannot be relied upon to infer the holdings of individual blockholders (Anderson and Lee (1997a and 1997b), Jeng et al. (2003)). Thus, we focus in this paper on the annual proxy data, which is more reliable and more commonly used by researchers.

Proxy data is available from many sources, including direct electronic access using the SEC's "Edgar" tool for all corporate filings since the mid-1990s. For large-scale data downloads, however, it is necessary to use a commercial product. The most widely used product is the *Compact Disclosure (CD)* database of Standard & Poor's. Anderson and Lee (1997a and 1997b) focus their analysis on the holdings of corporate officers and directors, and show that *CD* accurately reproduces the information in proxy statements for all firms except those with multiple classes of stock. While *CD* also reproduces data on blockholders from the tables in the proxy statement, there are additional problems with these data. We discuss these problems and their solutions in Section II, and summarize the changes for a large sample of firms from 1996 to 2001. For researchers using blockholder data in regression analysis, the raw data present an errors-in-variables problem. If blockholder data is used as a dependent variable, then these errors only cause biases if they are correlated with

the regressors. In section II we demonstrate that the errors are independent of a set of logical regressors, so bias is unlikely for many applications.

If, however, blockholders are used as an independent variable, then there are several possible biases. In Section III, we perform a representative study using both raw *CD* data and a “clean” data set where the *CD* data problems have been fixed. In our sample, we find that the raw data is much noisier: in annual regressions of Tobin’s *Q* on outside blockholder ownership and other control variables, the clean data set is far more likely to yield statistically significant point estimates for the ownership variables. Furthermore, bootstrap estimations demonstrate that improved precision is the typical outcome for this regression. The good news is that the bias appears restricted to the blockholder coefficients only, with no bias induced for the coefficients in other regressions.

Since our cleaned data is only available for a subset of firms and years, researchers will also be interested in the efficacy of alternative fixes for these data errors. In Section IV we discuss several alternatives based on truncating, winsorizing, or partial cleaning. While several of these fixes can alleviate the errors-in-variables bias, an economically significant bias still remains in all cases, with the best fix eliminating approximately one-half of the bias. Section V summarizes and concludes. Two appendices supplement the text. Appendix A provides details on the 1934 SEA and the disclosure requirements it created, and Appendix B provides details on the construction of our sample.

II. Data

A. Sample Firms

Our initial sample of firms consists of firms that are covered by the Investor Responsibility Research Center (IRRC) for both their publication *Corporate Takeover Defenses* (Rosenbaum 1995, 1998, 2000) and their director's database which provides details on the board of directors for about 1,500 of the largest U.S. companies. The IRRC's universe is drawn from the Standard & Poor's (S&P) 500 as well as the annual lists of the largest corporations in the publications of *Fortune*, *Forbes*, and *Businessweek*. We use the IRRC sample as a starting point because a wide range of governance data is available for this group of companies and our goal is to make this set of data as comprehensive as possible for this group.³ A special subset of the IRRC companies – less than 10 percent in all years – have multiple classes of common stock. For these companies, Anderson and Lee (1997a) showed there are many problems with the *CD* data, and these problems are very difficult to fix. In this paper, we eliminate all multiple-class companies from the database and start with the approximately 1300 firms per year (7,873 firm-years) for the single-classed companies in the IRRC sample from 1996 to 2001.⁴

The initial ownership data comes from the *CD* compact-disk product. Based on the results of Anderson and Lee (1997a, 1997b) we build our sample from the information on large shareholders that *CD* derives directly from the proxies and ignore the insider-trading data that is also available on the disks. Appendix B provides details on the construction of the initial database.

³ For example, the IRRC data has been used as a starting point by Gompers et al. (2003), Cremers and Nair (2004), and Gillan et al. (2003).

⁴ The dual-class companies are analyzed in a companion paper, Gompers et al. (2004), where we attempt to build a comprehensive sample of all dual-class companies with any share-class trading on any major exchange in the United States.

We next check the initial database by comparing the *CD* data to the original proxy statements, which we obtain from *Livedgar*,⁵ making changes to the ownership percentages of large shareholders where appropriate. All firms in the sample were checked – even those with no reported blockholders in *CD*. We employ the following general rules when deciding on share ownership. The SEC defines beneficial ownership as either voting or investment power, and sometimes companies report both measures in their proxies. We use voting power as opposed to investment power for our database when a distinction is made between the two. Also, even if individuals disclaim beneficial ownership of some portion of their holdings in the proxy, we treat these holdings as if the individual had the voting power. Under the terms of SEC Rule 13d-3, shares of common stock that may be acquired within 60 days are deemed outstanding for the purposes of computing the percentage of common stock owned by a shareholder. We follow this SEC rule and include these options. In the rare cases of a company having a temporary ownership structure resulting from a recent merger or acquisition, we remove these companies from our sample for that year. For 229 (2.9%) sample firms, proxy information could not be obtained, and these firm-years were removed from our sample. Our final sample consists thus of 7,649 firm-years and covers 1,913 unique firms. Table 1 shows summary statistics of our sample firms. The table is based on cross-sectional averages of time-series means.

Many researchers are interested in knowing whether a specific blockholder is an “insider” or an “outsider” to the firm. The role of a large shareholder in corporate governance is often treated differently depending on the classification of the shareholder. Since our work required the examination of all blockholders, the marginal cost of coding

⁵ *Livedgar* is an online data service, provided by Global Securities Information, Inc., that enables users to obtain source documents as filed with the SEC.

these classifications was relatively low, so we did so. The results are summarized in Table 2. The possible classifications are (1) officer, (2) director, (3) affiliated entity, (4) ESOP, and (5) outside blockholder. Category (1) includes all officers, even if they are also directors. Category (2) only includes non-officer directors. Category (3) includes any individual, trust, or company whose voting outcome is partially influenced, but not completely controlled, by an officer or director of the company. If the shares are completely controlled by the officer or director, then these shares would be counted under category (1) or (2), respectively. Category (4) is the aggregate number of shares held by Employee Share Ownership Plans, but does not include employee shares held through non-ESOP retirement plans (such as non-ESOP 401(k) plans). Category (5) includes all blockholders not elsewhere classified. This final category makes up about two-thirds of the aggregate amount of blockholding, and will be examined in the analysis of Section III.

B. Problems with the *Compact Disclosure* Data

Two main biases are introduced if researchers were to work directly with the raw *CD* database: **overlaps** and **preferred shares**. The SEC requires that all beneficial owners of more than 5% of a company's common stock be listed in the proxy, and consequently shares are often double or triple counted under different people or entities.⁶ While the SEC requires firms to detail the ownership structure of jointly held blocks in the footnotes, *CD* ignores all of the footnotes detailing joint or cross ownership of shares and lists every blockholder and ownership percentage exactly as it appears in the summary table of the proxy section "Security Ownership of Management and Certain Beneficial Owners." This leads to the overlap of reported ownership, which might be either a full overlap or a partial overlap.

⁶ See Appendix A for details of these disclosure requirements.

Examples of these two cases are documented below in Subsection 1. Second, *CD* sometimes misrepresents preferred shares as common equity ownership. This problem is illustrated below in Subsection 2. Subsection 3 discusses miscellaneous other problems.

1. Overlaps

Full overlaps can arise in two types of situations. In the first scenario, two or more blockholders are listed in the ownership table with the same shareholdings and the joint ownership of these shares is disclosed only by the footnotes. In the second scenario, the proxy separates the beneficial ownership of directors and officers from that of large shareholders and *CD* reproduces entries from both tables without cross-checking identities. Figures 1 and 2 display an example of the latter case. Figure 1 shows the *CD* data for Coca Cola Co. from the October 1999 disk, and Figure 2 shows an excerpt of the proxy statement from March 4th, 1999 on which the data is based. While *CD*'s blockholder data is accurate in a literal sense with respect to the two tables in the proxy, the vital information of the proxy footnote is ignored. Figure 1 lists Berkshire Hathaway and Warren Buffett individually as 8.10% owners of the common stock. Referring to the footnote 4 of Figure 1 (the ownership table from the proxy statement), we find that all of the shares listed under Warren Buffett are owned indirectly through Berkshire Hathaway. Tallying the beneficial ownership percentages without referencing the table footnotes in the proxy would suggest that 22.3% of Coca Cola's common stock is held by blockholders when actually 14.2% is the correct figure.

Overlaps are not always easy to recognize. In the second and more common scenario, multiple blockholders have joint-ownership of stock, but also own shares over which they have sole voting or investment power. This type of overlap does not result in identical

ownership figures for multiple shareholders. For example, there is no obvious overlap in the *CD* listing for Outback Steakhouse, Inc. (Figure 3). However, footnotes 1, 2 and 4 to the proxy beneficial ownership table (Figure 4) reveal that 8,242,125 shares (16.85%) owned by Multi-Venture Partners, Ltd. are also listed under Chris T. Sullivan, Robert D. Basham, and J. Timothy Gannon. This means that beneficial owners hold 26.98% of the common stock, rather than 77.53%, which is the straightforward sum of *CD*'s ownership statistics. Note that the Spectrum data does not report the 5% blockholders correctly. At the bottom of Figure 3, where the Spectrum data is listed after the delimiter '****^', it is stated that there are five 5% owners who hold an aggregate of 62.06% of common stock.

Both of the above examples are relatively easy to spot and correct. However, many companies have more complex overlaps which translate into more detailed footnotes and longer amounts of time that a researcher must spend examining them. These types of overlaps include those among companies, subsidiaries, individuals, and trusts. When the information in the footnotes is insufficient to determine the ultimate control of these shares, we follow the rule of assigning these shares to the partial owner who is closest to control of the company: officers first, then directors, and then outsiders.

2) Preferred Stock

There are two ways in which preferred shares can erroneously enter into the beneficial ownership figures. A company may report ownership of common stock and preferred stock separately but side-by-side in the same proxy table. In this case, *CD* will pick up all of the percentages in the table without distinguishing between the two categories of stock, giving us two different ownership figures for shareholders holding both preferred and

common stock. Figure 5 shows one example from the original *CD* database, and Figure 6 shows the corresponding proxy statement. Some companies structure their proxy statements by listing common and preferred ownership side-by-side, but also giving an aggregated ownership figure. In these cases, *CD* tends to erroneously pick up just the aggregate ownership figure, without doing any adjustment for the relative voting power between the common and preferred. In our corrected data, we include only the common-stock component of voting.

3) Other problems

The vast majority of other problems are blocks that are simply missed by *CD*. Most often, this stems from the firm not following the standard procedure of having a table with the names of certain beneficial owners or principal shareholders, but rather embedding this information in the text. *CD* does not capture information from the text.

In addition, *CD* sometimes picks up information from both a preliminary proxy and a final proxy. Each shareholder would then be recorded twice, with two different proxy source dates and the same percentage of holdings. Also, in rare cases, *CD* makes typos entering the percentage of shares held (i.e. 9.08% instead of 9.80%). Finally, sometimes the proxy allows us to break out blockholdings more specifically. Most often, it would be a bank that holds a certain amount of shares in its capacity as the ESOP trustee and another amount for its retail customers. Often, this number is lumped together in the proxy, and if the footnote allows us to separate them, we do.

C. The Corrected Data

The first four rows of Table 3 show the frequency of the corrections we made by groups of overall shares held by blockholders. Using a non-parametric Wilcoxon difference test, we find that the corrected and raw percentages of rows three and four are statistically different for all 6 groups of blockholders at the one percent level. The remainder of Table 3 shows the frequency of each of the problems. It is evident from Table 3 that the overlap problem becomes more pronounced the higher is the overall share of common stock held by blockholders. For companies where the reported ownership in *CD* is between five and ten percent, the frequency of errors is 0.4 percent. This error rate rises to 4.5 percent for the 10-15 percent ownership range, 8.5 percent for the 15-25 percent range, 13.5 percent for the 25-50 percent range, and 53.1 percent for the >50% range. The category with no reported blockholders in *CD* is a special case: there, the 31.7 percent of errors is caused by *CD* erroneously reporting no blockholder.

Errors tend to increase with aggregate blockholdings in *CD* for the simple reason that the errors themselves tend to increase the aggregates: e.g., an overlap for a jointly held block of 25 percent will lead to a *CD* aggregate of 50 percent. While the errors are rare for low levels of aggregate holdings, they are common at higher levels and are economically significant. In the most extreme category (>50% in *CD*), the average holdings in the raw data for firms with errors fall from over 100 percent (clearly impossible) to under 50 percent.

Empirical economists are accustomed to working with noisy data. The impact of measurement error often depends upon the correlation of the noise with other explanatory variables. Table 4 shows the correlations among the blockholder variables, including the error term (clean blockholder holdings minus raw blockholder holdings) and several other

firm-level characteristics. The good news for researchers is that the error term appears to be uncorrelated with all of these variables. Thus, regressions with blockholder data as the dependent variable are likely to provide unbiased coefficient estimates – at least for these variables. Of course, the additional noise in the dependent variable will tend to increase standard errors, but with enough data this problem can be overcome.

Table 5 summarizes the output for a multivariate Tobit regression of outside blockholder holdings on the logarithms of firm age and the book value of assets. These two regressors are chosen because they are the only two characteristics in Table 4 that are significantly correlated with blockholding. The Tobit regression is left-censored at 0. The table shows no clear bias in the coefficients for either regressor, but the standard errors are always higher for the raw regressions. Nevertheless, there is sufficient data so that differences for inference are minor.

III. A Representative Analysis: Outside Blockholders and Firm Value

If blockholder data is used as an independent variable, then there are several potential sources of bias. This measurement error (“errors-in-variables”) problem is well-studied by econometricians, with the quantitative importance of the problem depending upon the severity of the measurement error and the correlation of this error with other variables of interest.⁷ In this section, we assess the impact of this problem using a representative analysis of the relationship between firm value and outside blockholdings. Specifically, for each year t , we estimate

$$Q_{it} = a + b_1 OWN_{it} + b_2 OWN_{it}^2 + cX_{it} + e_{it}, \quad (1)$$

⁷ For a textbook treatment of the problem, see Greene (1997), p.435-444.

where Q is a measure of industry-adjusted Tobin's Q , OWN and OWN^2 are the fraction and fraction squared held by outside blockholders, X is a vector of control variables, and e is an iid error term, all measured for firm i at time t . We follow Kaplan and Zingales (1997) and measure Q as the ratio of the market value of assets to the book value of assets: the market value is calculated as the sum of the book value of assets and the market value of common stock less the book value of common stock and deferred taxes. The X vector includes a Delaware incorporation dummy (Daines, 2001), an S&P 500 inclusion dummy (Mørck and Yang, 2001), the log of sales for the previous fiscal year, and the log of firm age (months since first public listing).

We make no claims for any causal relationship here. Any regression of firm value on ownership will be fraught with endogeneity concerns, a point first made by Demsetz and Lehn (1985) and followed up by many other authors.⁸ Rather, we intend only to test whether all our cleaning effort yields any meaningful differences in coefficient estimates or statistical significance. To do so, we start with this simple and obvious regression.

Table 6 summarizes the results of regression (1) estimated separately for each year from 1996 to 2001, with each regression estimated using both raw and cleaned data. The results suggest that the raw data is much noisier, as the coefficients on the OWN variables are almost always closer to zero and lower in statistical significance than are the corresponding coefficients using cleaned data. Overall, the cleaned data demonstrate a far more robust relationship between outside block ownership and firm value. While the economic interpretation of this result is clouded by endogeneity concerns, the research importance of using the cleaned data is clear.

⁸ For examples, see Loderer and Martin (1997), Himmelberg et al. (1999), Palia (2001), and Coles et al. (2003).

The results of Table 6 demonstrate that our proposed corrections matter for one “draw” of history. To get a better understanding for the robustness of these results, we use a bootstrap method to simulate 25,000 draws. Instead of drawing from a specified distribution, the bootstrap draws with replacement from the sample. For a linear regression model like (1), the sample provides the empirical distribution for the dependent variable, the independent variables and the error term as well as estimates for constant, slope, and error variance.

Our example focuses on one year (1998) of data only, because it is reasonable to assume that the assumptions for an iid nonparametric bootstrap with paired sampling are fulfilled for cross-sectional data. Time-series and panel data would require more sophisticated bootstrap techniques without a corresponding advantage for our analysis.

Specifically, we begin with the 1998 observed sample of $n = 1, \dots, 1,216$ independent observations for our dependent variable ($\mathbf{y} = \mathbf{y}_1, \dots, \mathbf{y}_{1216}$) (industry-adjusted Q), and the independent variables ($\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_{1216}$.) of the regression in Table 6. A paired bootstrap sample is obtained by independently drawing $N = 25,000$ pairs $(\mathbf{x}_i, \mathbf{y}_i)$ from the observed sample with replacement. While the bootstrap sample has the same number of observations, some observations appear several times and others never. We then estimate (1) with each of the 25,000 bootstrapped samples.

Figures 7 and 8 illustrate the results for the coefficients on outside blockholders and outside blockholders squared, respectively. It is clear from the figures that the results of Table 6 are no fluke, as bootstrap distributions appear quite different in the two cases (raw and cleaned). These figures illustrate the empirical distribution of the coefficients across all 25,000 draws. If we compare the coefficients within the same draw, we find that the absolute coefficient on blockholders for the cleaned data is greater than its corresponding coefficient

for the raw data in 97.6% of the 25,000 draws. Similarly, the coefficient on blockholders squared for the cleaned data is greater than its corresponding coefficient for the raw data in 94.4% of the 25,000 draws. For the coefficient on blockholdings, 86.5% of the cleaned-sample estimates are negative and statistically significant at the 5% level, while only 43.6% of the raw-sample estimates meet those conditions. For the coefficient on blockholdings squared, 52.4% of the cleaned-sample estimates are positive and statistically significant, while only 15.9% of the truncated-sample estimates meet those conditions. Overall, we conclude that measurement error is a significant problem for inference on these two coefficients.

Measurement error for one regressor (blockholders) may also affect inference for the other regressors. There are two concerns here. First, if the measurement error (for blockholders) is correlated with other regressors, then the coefficients on these regressors would themselves be subject to measurement-error bias. Second, as measurement error (for blockholders) goes to infinity, then the blockholder regressor is effectively omitted from the estimation and the other regressors are subject to omitted-variable bias, with the direction of this bias dependent upon the full set of covariances.

The evidence of Tables 3 and 4 in Section II suggests that the first problem is not quantitatively important: since the blockholder error appears uncorrelated with the other regressors, there is no direct bias induced by this error. To quantify the importance of the second problem, Table 7 summarizes the bootstrap estimates for the other coefficient estimates in regression (1), using the same draws as summarized in Figures 7 and 8. The table summarizes analogues to a “95%-confidence interval” around the median draw. When comparing the corresponding intervals from the raw and cleaned data, there is no apparent

pattern, and certainly no significant pattern. Thus, we conclude that the measurement error does not contaminate other coefficients in (1).

IV. Alternatives to Cleaning the Whole Sample?

It may not always be practical to use a cleaned sample of blockholdings. For an analysis outside the sample or time period discussed in this paper, researchers may still need to use raw data. Thus, it may be helpful to understand the efficacy of alternative (cheaper) fixes for the raw *CD* data. Since we found economically significant bias only when blockholders were used as an independent variable, we focus attention on that case in this section, and we repeat OLS regression (1) for 1998 using several alternatives: samples truncated at 50% or 100% blockholder ownership; samples winsorized at 50% or 100% blockholder ownership; and samples cleaned only for firms with blockholder ownership greater than 50% or 100%. We also estimate a median regression of (1). The use of 50% and 100% is arbitrary but seems reasonable for this application, and also is close to the standard cutoffs of 95th percentile (which would be 58% ownership here) and 99th percentile (which would be 108% ownership here), respectively.

Table 8 reports the results. It does not appear that any of these low-cost fixes are effective. While all truncations or winsorizations appear to reduce the errors-in-variables bias, none of these simple fixes would yield inferences similar to those of the cleaned regression in Table 4. The best results from this group are obtained by truncating at 100% blockholdings, or by manually cleaning all blocks over 50% ownership or over 100% ownership. All three of these best alternative fixes yield about the same changes in the key

coefficients. Since truncation at 100% is certainly the least costly fix to implement, it seems to be the alternative of choice.

To further investigate the use of truncation at 100%, we also obtain 25,000 bootstrap estimates using the same procedures as in Section III. These bootstrap estimates are consistent with the findings of Table 8. For the coefficient on blockholdings, 86.5% of the cleaned-sample estimates are negative and statistically significant at the 5% level, while only 55.6% of the truncated-sample estimates meet those conditions. For the coefficient on blockholdings squared, 52.4% of the cleaned-sample estimates are positive and statistically significant, while only 35.3% of the truncated-sample estimates meet those conditions.

V. Conclusion

Researchers rely on ownership data for many studies. The lack of a standardized source of data on large blockholders is an impediment to this work. In this paper, we document the weaknesses with the commonly used data, show how to fix them, and demonstrate that these fixes are both quantitatively large and also important for some applications.

The measurement error in blockholder data creates several possibilities for bias. Our analysis suggests that empirical work with blockholder data as the dependent variable will produce unbiased results, as the measurement error is not correlated with the other firm-level characteristics that we tested. While the measurement error does increase standard errors, the increase is not severe.

Researchers who use blockholder data as an independent variable face a larger challenge. In a representative analysis of firm value and blockholdings, we find that using

the uncorrected raw data leads to significant biases for the blockholder coefficients, and simple fixes such as truncating or winsorizing the sample provide only partial alleviation of these biases. The coefficients on other regressors do not appear to have biases. Thus, if the blockholder effects are the key independent variable, we believe it is necessary to work with a cleaned sample. If blockholder data is only being used as a control variable, then a cleaned sample is much less crucial.

Appendix A – Legal Rules

The Securities Exchange Act of 1934, rules 13d-1 to 13d-7 (§240.13d) and rules 16a-1 to 16e-1 (§240.16a-e), contain the legal definitions and filing requirements for what the corporate finance literature has termed “blockholders”. While 13D defines beneficial ownership and describes the disclosure requirements of these shareholders, rules 16a-e detail the reporting requirements on transactions made by corporate insiders. The company in turn is required to disclose blockholder information to shareholders via proxies under Regulation and Schedule 14a (§240.14a), commonly called the “proxy” statement.

Rule 13d-1(a) sets the threshold for beneficial ownership at 5% or more of a class of stock. According to Rule 13d-3(a) a beneficial owner “includes any person who, directly or indirectly, through any contract, understanding, relationship, or otherwise, has or shares” voting or investment power. This rule has been interpreted to include shares that may be obtained through the exercising of options, warrants, or rights in the next 60 days a part of the beneficial ownership calculation. Any individual or group that has acquired a beneficial stake in a class of equity is required to file the form SC 13D [see rule 13d-1(a)]. According to Rule 13d-3(a), the form SC 13D must be filed within 10 days after the acquisition of the

equity position, and any material change in the position must be promptly filed in an amended SC 13D [see 13d-1(a)]⁹. This form contains information about the person's relationship with the company and the nature of the holding. A select category of "persons" such as banks, brokers and dealers, and insurance companies can file an abbreviated form, the SC 13G. The SC 13G can only be used if the equity securities were acquired in the ordinary course of business, and without the purpose or effect of changing or influencing the control of the issuer [see 13d-1(b)].

Rule 13d-3 details how to determine beneficial ownership. For the purposes of sections 13(d) and 13(g) of the Act, a beneficial owner of a security includes any person who, directly or indirectly has or shares either voting or investment power over the security. Rule 13d-3 also specifies that creating trusts or pooling arrangements are not valid constructs to avoid filing a form 13D. In fact, rule 13d-3(c) specifies, "all securities of the same class beneficially owned by a person, regardless of the form which such beneficial ownership takes, shall be aggregated in calculating the number of shares beneficially owned by such person". Rule 13d-7 regulates the dissemination of form 13D. For our purposes, it is important that the issuer of the security must be notified of the existence of a large shareholder at its principal executive office, by registered or certified mail.

The second set of laws of the Securities Exchange Act of 1934 regulating filing requirements for blockholders is detailed in rules 16a-1 to 16e-1 ("Reports of Directors, Officers, and Principal Shareholders"). Pursuant to this section, a person ("principal shareholder") deemed a beneficial owner of more than 10% of any class of equity securities registered under section 12 of the Act ("Registration Requirements for Securities"), as well as any director or officer of the company needs to file forms 3, 4, and 5. Initial statements of

⁹ According to rule 13d-2(a), The SEC deems a 1% change in the ownership position as material.

beneficial ownership of equity securities required by section 16(a) of the Act are filed on Form 3. Statements of changes in beneficial ownership required by that section are filed on Form 4, and annual statements are filed on Form 5. While the 13D/G forms are laid out to disclose the size and nature of the holdings of a large blockholder (who owns more than five percent, what their purpose is, whether they intend to buy more, their voting power, etc.), the 3,4,5 forms are structured to disclose any transactions made by a company insider. Without discussing anything about the intent of the shareholder, the 3,4,5 forms list what was bought or sold and when, making them more transaction oriented than disclosure oriented.

The laws regulating a company's disclosure requirements of large blockholder information to shareholders are detailed in Regulation 14A ("Solicitation of Proxies") and Schedule 14A of the Securities Exchange Act of 1934 (§240.14a). Pursuant to Schedule 14A(6-d) with reference to Item 403 of Regulation S-K (§229.403) entitled "Security Ownership of Certain Beneficial Owners and Management," a company is required to disclose two types of tables in proxies: one listing shareholders who own more than 5% of any class of the company's equity securities, and another listing any ownership of equity securities by all directors and officers of the company. The tables can be combined at the company's option and should list the number of shares beneficially owned and the percentage of the class owned. Item 403 does specify a tabular format to be used, but it does not specify the location where this information should appear in the proxy. Item 403 also explains that a company may rely on the information disclosed in the SC 13D/G forms by beneficial owners when preparing this information for proxies.

Item 403 refers to Rule 13d-3 for the determination of beneficial ownership, and it details specific guidelines for disclosing the nature of the beneficial ownership. Information

in sub-columns or footnotes must be included and reflect the number of shares of which a beneficial owner has sole or shared voting power as well as sole or shared investment power. Similarly, the number of shares a beneficial owner has the right to acquire must be listed in a footnote, as well as any arrangements or pledges that could lead to a change of control of shares. Item 403 also specifies that a company must use appropriate disclosure to avoid confusion where more than one beneficial owner is listed for the same securities. When a beneficial owner owns shares pursuant to a voting trust or agreement, the company must state in a table or footnote the title of the securities, the amount held or to be held according to the trust or agreement, and the duration of the agreement. The company must also disclose the names and addresses of the voting trustees and outline briefly their voting rights and powers under the trust or agreement.

Appendix B – Details on the Sample Construction

In a first step, we matched the IRRC database with data from *CD* pertaining to large shareholders and directors and officers. *CD* stores company data on compact disks that are updated monthly. The month of the update depends on the company's fiscal year end, and, for data pertaining to the board of directors, on the company's proxy meeting date. Since *CD* often does not keep information until the next update, but rather removes stale data from the monthly disks, searching the right month after the proxy meeting is important. In the earlier sample years, a lag of 6 months from proxy meeting month to issuance of data disk yielded the most reliable results. Starting in 2000, *CD* data are usually updated in the month following the annual meeting.

We then sort, for each sample year, all IRRC firms by proxy meeting month. We download through a ticker merge the large shareholder data (field SH) and information on directors (field DO) and officers (field NA) from the appropriate monthly *CD* disk in the tagged format. Figure 1 shows an example of our raw database. The search string automatically yields a summary of the insider-trading data compiled by *Spectrum* (Form 3, 4, and 5), which we discard.¹⁰ Note from Figure 1 that *CD* mentions the source and source date of the director and ownership information in parentheses. Using this information, we ensure that all ownership information is taken from proxy statements. We are able to match 94% of all firm-years with our search strategy through *CD*. The remaining firms were looked up directly from the proxies. Thus, 6% of the final *CD* sample is comprised of entries that we constructed ourselves from the proxies.

In a next step, we use a *SAS* program to convert the data of Figure 1 into an easy-to-use *SAS* database. The *SAS* program loops through the information on officers and directors, provided by *CD* in the fields NA- and DO-, and identifies a large shareholder entry as either director or officer if the character string for a blockholder (in the SH- section of the *CD* data) matched a character string in the officer or director sections (NA-, DO-) of the data. *CD* is extremely accurate and consistent in the spelling of names across fields, as it takes the names exactly as they appear in the proxy statements.¹¹

¹⁰ Anderson and Lee (1997a) caution that the proxy's definition of an insider (an officer or director) is not the same as the definition used for the insider-trading filings (which also includes any shareholder of 10% or more) compiled by *Spectrum*. When they compare *Spectrum* data to the benchmark proxy data on management ownership, they find that 40.7% of their sample has a reporting discrepancy of over 5% (p. 316). They also refute the claim that the *Spectrum* data are more current than the most recent proxy data, showing that the filing dates listed in the *Spectrum* section of the *CD* data are sometimes "stale" by two or more years (Anderson and Lee (1997b), p. 3-4). An additional concern for us is a date mismatch, as the *Spectrum* data is updated at various points during the year, while the director's database stems from the annual proxy statements. Indeed, the *Spectrum* data of Figure 1 is incorrect.

¹¹ However, the automated process will not correctly match variations such as "The Smith family trust" with a director named Peter H. Smith. These corrections are done manually in the data cleaning process.

References

Anderson, Ronald C., Thomas W. Bates, John M. Bizjak, and Michael L. Lemmon, 2000, Corporate Governance and Firm Diversification, *Financial Management* 29, 5-22.

Anderson, Ronald C. and D. Scott Lee, 1997a, Ownership Studies: The Data Source Does Matter, *Journal of Financial and Quantitative Analysis* 32, 311-329.

Anderson, Ronald C. and D. Scott Lee, 1997b, Field Guide for Research Using Ownership Data, *Journal of Financial and Quantitative Analysis* 32 (supplement), 1-11.

Ang, James S., Rebel A. Cole, and James Wuh Lin, 2000, Agency Costs and Ownership Structure, *Journal of Finance* 55, 81-106.

Coles, Jeffrey L., Michael L. Lemmon, and J. Felix Meschke, 2003, Structural Models and Endogeneity in Corporate Finance: The Link Between Managerial Ownership and Corporate Performance, Arizona State University Working Paper.

Cremers, Martijn and Vinay B. Nair, 2004, Governance Mechanisms and Equity Prices, forthcoming *Journal of Finance*.

Daines, Robert, 2001, Does Delaware Law Improve Firm Value?, *Journal of Financial Economics* 62, 525-558.

Demsetz, Harold and Kenneth Lehn, 1985, The Structure of Corporate Ownership: Causes and Consequences, *The Journal of Political Economy* 93, 1155-1177.

Denis, David J., Diane Denis, and Atulya Sarin, 1997, Ownership Structure and Top Executive Turnover, *Journal of Financial Economics* 45, 193-221.

Fama, Eugene F., and Kenneth R. French, 1997, Industry Costs of Equity, *Journal of Financial Economics* 43, 153-194.

Gillan, Stuart L., Jay C. Hartzell, and Laura T. Starks, 2003, Industries, Investment Opportunities and Corporate Governance Structures, University of Delaware John L. Weinberg Center for Corporate Governance Working Paper.

Gompers, Paul A., Joy L. Ishii, and Andrew Metrick, 2003, Corporate Governance and Equity Prices, *Quarterly Journal of Economics* 118, 107-155.

Gompers, Paul A., Joy L. Ishii, and Andrew Metrick, 2004, Incentives vs. Control: An Analysis of U.S. Dual-Class Companies, NBER WP #10240.

Greene, William H., 1997, *Econometric Analysis*, 3rd Edition, Prentice Hall, New Jersey.

Heflin, Frank and Kenneth Shaw, 2000, Blockholder Ownership and Market Liquidity, *Journal of Financial and Quantitative Analysis* 35, 621-633.

Himmelberg, Charles P., R. Glenn Hubbard, and Darius Palia, 1999, Understanding the Determinants of Managerial Ownership and the Link Between Ownership and Performance, *Journal of Financial Economics* 53, 353-384.

Holderness, Clifford G., 2003, A Survey of Blockholders and Corporate Control, *FRBNY Economic Policy Review*, April 2003.

Jeng, Leslie A., Andrew Metrick, and Richard Zeckhauser, 2003, Estimating the Returns to Insider Trading: A Performance-Evaluation Perspective, *Review of Economics and Statistics* 85, 453-471.

Kaplan, Steven N., and Luigi Zingales, 1997, Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?, *Quarterly Journal of Economics* 112, 169-216.

Loderer, Claudio, and Kenneth Martin, 1997, Executive Stock Ownership and Performance: Tracking Faint Traces, *Journal of Financial Economics* 45, 223-255.

McConnell, John J., and Henri Servaes, 1990, Additional Evidence on Equity Ownership and Corporate Value, *Journal of Financial Economics* 27, 595-612.

Mørck, Randall K., Andre Shleifer and Robert Vishny, 1988, Management Ownership and Market Valuation: An Empirical Analysis, *Journal of Financial Economics* 20, 293-315.

Mørck, Randall K., and Fan Yang, 2002, The Mysterious Growing Value of the S&P 500 Membership, University of Alberta Working Paper.

Palia, Darius, 2001, The Endogeneity of Managerial Compensation in Firm Valuation: A Solution, *Review of Financial Studies* 14, 735-764.

Rosenbaum, Virginia, 1995, 1998, 2000, *Corporate Takeover Defenses*, Investor Responsibility Research Center Inc., Washington, D.C.

Ryan, Harley E., Jr. and Roy A. Wiggins III, 2001, The Influence of Firm- and Manager-Specific Characteristics on the Structure of Executive Compensation, *Journal of Corporate Finance* 7, 101-123.

Shivdasani, Anil, 1993, Board Composition, Ownership Structure, and Hostile Takeovers. *Journal of Accounting and Economics* 16, 167-198.

Singh, Manohar and Wallace N. Davidson III, 2003, Agency Costs, Ownership Structure and Corporate Governance Mechanisms, *Journal of Banking and Finance* 27, 793-816.

Table 1
Sample Summary Statistics

	<u>Mean</u>	<u>Median</u>
Firm Characteristics		
Firm Age (in years)	23	18
Market Value	5,966	1,216
Assets - Total (MM\$)	8,843	1,374
Sales (Net) (MM\$)	3,601	1,088
Common Equity – Total (MM\$)	1,586	477
Book-to-Market Ratio	0.61	0.50
Capex-to-Sales Ratio	0.15	0.05
Valuation and Performance		
Tobin's Q	1.97	1.41
Annualized Stock Return	-1.0%	6.3%
Return on Assets	3.8%	4.8%
Return on Equity	11.6%	11.2%
Governance Characteristics		
Governance Index	9.0	9.0
Number of Directors	9.5	9.0
% Independent Directors	61.2%	63.6%
% Dependent Directors	38.8%	36.4%
% Employee Directors	22.2%	19.4%

The table describes summary statistics of key sample firm characteristics. The sample period is 1996 – 2001. The total sample consists of large publicly listed U.S. firms, a total of 7,649 firm-years and 1,913 unique firms. The table contains cross-sectional means and medians of firm time-series averages. Firm age is measured as months since first listing. Market value, total assets, sales, and common equity are taken from Compustat. The book-to-market ratio is book value of common equity to market value of common equity. Book value of common equity is the sum of book common equity (Compustat item 60) and deferred taxes (item 74). The capex-to-sales ratio is capital expenditures (item 128) divided by net sales. Q is the market value of assets divided by the book value of assets (item 6), where the market value of assets is computed as book value of assets plus the market value of common stock less the sum of the book value of common stock (item 60) and balance sheet deferred taxes (item 74). All book values for fiscal year t (from Compustat) are combined with the market value of common equity at the calendar end of year t. Return on assets (equity) is calculated as income before extraordinary items (item 18) divided by item 6 (item 60). The Governance Index is a measure of shareholder rights developed by Gompers, Ishhii, and Metrick (2003) and available through Wharton Research Data Services (WRDS). All director variables are from the director database of the Investor Responsibility Research Center (IRRC).

Table 2
Affiliation of Large Shareholders

	1996	1997	1998	1999	2000	2001
Total Firms in Sample	1,130	1,046	1,510	1,387	1,336	1,240
Number of Blockholders	2.12	2.10	2.41	2.44	2.53	2.50
Sum of Blockholdings (%)	21.7%	21.3%	24.5%	24.9%	25.5%	25.0%
Held by						
Officers						
<i>Number</i>	0.16	0.15	0.20	0.20	0.21	0.20
<i>% Held</i>	2.2%	2.1%	2.7%	2.7%	2.8%	2.5%
Directors						
<i>Number</i>	0.09	0.10	0.11	0.12	0.12	0.11
<i>% Held</i>	1.1%	1.1%	1.3%	1.4%	1.5%	1.3%
Affiliated entities						
<i>Number</i>	0.14	0.14	0.13	0.14	0.13	0.13
<i>% Held</i>	2.6%	2.5%	2.3%	2.2%	2.0%	2.1%
ESOPs						
<i>Number</i>	0.12	0.13	0.09	0.09	0.09	0.10
<i>% Held</i>	1.3%	1.3%	0.9%	0.9%	1.0%	1.0%
Outside Blockholders						
<i>Number</i>	1.60	1.58	1.87	1.90	1.97	1.97
<i>% Held</i>	14.4%	14.3%	17.2%	17.7%	18.2%	18.0%

The first three rows of this table contain summary statistics for the cleaned sample by calendar year. The sample construction is explained in Appendix B. The bottom part of the table shows the frequency of each category of large shareholders. **Officers** are officers, and possibly also directors of the firm. **Directors** are all non-officer directors. **Affiliated entities** are individuals, trusts or companies whose voting outcome is at least partially influenced or outright determined by an officer or director of the company. **ESOPs** are Employee Stock Ownership Plans. **Outside Blockholders** are all individuals or entities that are none of the above.

Table 3**Occurrence of Full and Partial Overlap and Wrong Attribution of Preferred Shares**

	Fraction of Common Stock Held by Large Shareholders						Total
	no BH	5-10%	10-15%	15-25%	25-50%	>50%	
Overall Sample							
No. of Firms before cleaning	1,365	802	738	1,469	2,251	1,024	7,649
% of Firms that required cleaning	31.7%	0.4%	4.5%	8.5%	13.5%	53.1%	18.8%
% Held as Blocks before cleaning	0.0%	7.0%	12.6%	19.9%	35.4%	82.4%	27.2%
after cleaning	7.9%	6.9%	12.4%	19.5%	33.6%	52.6%	24.0%
Situations that required cleaning:							
Full Overlap							
	0	0	1	20	47	122	190
% Held as Blocks before cleaning			14.8%	21.3%	37.6%	99.9%	
after cleaning			6.9%	14.6%	24.2%	48.6%	
Partial Overlap							
	0	0	5	43	216	377	641
% Held as Blocks before cleaning			12.0%	21.6%	39.8%	97.3%	
after cleaning			6.2%	14.0%	27.4%	47.4%	
Preferred Shares							
	0	3	10	11	12	79	115
% Held as Blocks before cleaning		6.4%	12.6%	18.8%	34.1%	125.8%	
after cleaning		0.0%	3.2%	12.2%	11.5%	26.2%	
Other							
	433	0	22	25	24	22	526
% Held as Blocks before cleaning	0.0%		12.2%	19.8%	36.1%	85.4%	
after cleaning	24.7%		11.9%	19.0%	28.5%	42.2%	
Overall sample after cleaning							
Number of firms by category	963	952	854	1,640	2,518	722	7,649
% Held as Blocks by category	0%	7%	13%	20%	35%	62%	24.0%

This table reports the occurrence of the four problems associated with using the raw large shareholder database available through *Compact Disclosure*. We classify the firms into six different categories based on the sum of the blockholdings. The first row shows the number of firms by blockholding category. The second row shows the fraction of entries by category that needed to be corrected. The third and fourth row contain the average sum of blockholdings by category before and after cleaning, where the categorization is done by the sum of blockholdings as they appear in the raw data. The second part of the table describes the four situations that required cleaning. The four problems are full overlap, partial overlap, the treatment of preferred shares, and other problems. The second and third row of each reported problem show how the percentage holdings change given we observe the problem.

Table 4
Pearson Correlation Coefficients for 1998

	Sum of blockholdings	Sum of outside bh	Sum of inside bh	Bh Error	log (book va- lue assets)	log (firm age)	Capex / sales	Capital intensity	R&D intensity	Leverage	Cashflow / assets
Sum of blockholdings	1	0.74**	0.32**	-0.027	-0.29**	-0.30**	0.02	-0.05*	-0.09	0.007	0.005
Sum of outside bh		1	-0.13**	0.04	-0.22**	-0.26**	0.03	-0.03	0.04	0.04	-0.03
Sum of inside bh			1	-0.07**	-0.03	-0.03	0.01	0.003	-0.04	0.03	0.03
Bh_error				1	0.01	-0.01	0.008	0.01	0.005	-0.015	-0.018

The table shows the Pearson correlation coefficients for the sample year 1998. Significance at the 1% level is denoted by **, and significance at the 5% level is denoted by *. The variables are defined as follows. Sum of blockholdings is the sum of 5% blocks held by all entities from the cleaned database. Sum of outside (inside) bh is the sum of all blocks held by entities that were classified as outsiders (insiders or affiliates). Bh Error is the sum of 5% blocks held by all entities from our cleaned database minus the sum of 5% blocks held by all entities from the raw database. Book value of assets is taken from Compustat's industrial annual database, item 6. Firm age is the number of months since the first listing on a stock exchange. Capex / sales is item 128 divided by item 12. Capital intensity and R&D intensity are defined as in Himmelberg, Hubbard, and Palia (1999). Capital intensity is PPE (item 30) over sales. R&D intensity is R&D (item 46) over sales. Leverage is defined as long term and short term debt (item 9 and 34) over assets. Cashflow is defined as income before extraordinary items (item 18) plus depreciation and amortization (item 14)

Table 5**Annual Tobit Regressions of Outside Blockholder Holdings on Explanatory Variables**

	1996	1997	1998	1999	2000	2001
Raw Compact Disclosure Data						
Intercept	0.44** (0.07)	0.58** (0.07)	0.50** (0.06)	0.59** (0.05)	0.37** (0.06)	0.45** (0.06)
Log (book value assets)	-0.014* (0.006)	-0.023** (0.006)	-0.010 (0.006)	-0.017** (0.005)	-0.020** (0.006)	-0.028** (0.005)
Log (firm age in months)	-0.043** (0.012)	-0.049** (0.012)	-0.052** (0.010)	-0.049** (0.010)	-0.007 (0.011)	-0.015 (0.012)
Cleaned data						
Intercept	0.52** (0.05)	0.52** (0.05)	0.55** (0.04)	0.57** (0.04)	0.49** (0.04)	0.55** (0.04)
Log (book value assets)	-0.016** (0.004)	-0.019** (0.004)	-0.016** (0.004)	-0.017** (0.004)	-0.020** (0.004)	-0.033** (0.004)
Log (firm age in months)	-0.050** (0.009)	-0.047** (0.009)	-0.055** (0.007)	-0.052** (0.007)	-0.032** (0.007)	-0.026** (0.008)

This table presents the coefficients, before and after our cleaning procedure, from annual cross-sectional Tobit regressions of the fraction of the firm held by outside blocks on the logarithm of book value of assets and firm age. The natural logarithm of firm age in months is measured as months since first listing, and is obtained from CRSP. Book value of assets is obtained from Compustat (item 6). Standard errors are reported in parentheses and significance at the one-percent and five-percent levels is indicated by ** and * respectively.

Table 6
Q Regressions Using Raw *Compact Disclosure* and Our Cleaned Data

	1996	1997	1998	1999	2000	2001
Raw <i>Compact Disclosure</i> data						
Intercept	2.78** (0.40)	2.98** (0.44)	4.22** (0.50)	5.44** (0.75)	6.15** (0.63)	3.09** (0.31)
% held through outside blocks	-1.04* (0.44)	-1.10* (0.46)	-0.81 (0.46)	-1.85** (0.72)	-2.46** (0.70)	-0.96** (0.30)
% held through outside blocks squared	0.71 (0.60)	0.89 (0.58)	0.44 (0.34)	0.89 (0.62)	1.85* (0.72)	0.43 (0.28)
Cleaned data						
Intercept	2.95** (0.41)	3.19** (0.44)	4.59** (0.52)	5.94** (0.77)	6.73** (0.65)	3.42** (0.32)
% held through outside blocks	-1.45* (0.65)	-2.25** (0.70)	-2.69** (0.91)	-4.49** (1.33)	-4.78** (1.09)	-2.15** (0.54)
% held through outside blocks squared	0.96 (1.24)	2.33 (1.35)	3.29* (1.57)	5.60** (2.22)	5.75** (1.80)	1.98* (0.93)

This table presents the coefficients of the sum of outside blockholdings and squared outside blockholdings, before and after our cleaning procedure, from annual cross-sectional regressions of industry-adjusted Tobin's Q on the blockholdings and control variables. The control variables include a Delaware dummy, S&P 500 affiliation, the natural logarithm of firm age in months (measured as months since first listing, obtained from CRSP), and the natural logarithm of firm sales (obtained from Compustat). For simplicity, the coefficients on the control variables are omitted from the table. Standard errors are reported in parentheses and significance at the one-percent and five-percent levels is indicated by ** and *, respectively. Entries in **bold** denote a difference in the level of statistical significance between the coefficients in the regressions for the raw and the cleaned data. Q is the ratio of the market value of assets to the book value of assets: the market value is calculated as the sum of the book value of assets and the market value of common stock less the book value of common stock and deferred taxes. The market value of equity is measured at the end of the current calendar year, and the accounting variables are measured in the current fiscal year. Industry adjustments are made by subtracting the industry median, where medians are calculated by matching the four-digit SIC codes from December of each year to the 48 industries designated by Fama and French [1997].

Table 7**Bootstrapped Confidence Intervals for explanatory variables in Q regressions for 1998
for Raw Compact Disclosure and Cleaned Data**

		95% Confidence
Raw Data	Coefficient	Interval
log (book value of assets)	-0.176	[-0.257 , -0.094]
log (age)	-0.430	[-0.592 , -0.267]
delaware dummy	-0.048	[-0.273 , 0.177]
SP500 inclusion dummy	1.121	[0.779 , 1.462]
fraction held by outside blockholders	-0.806	[-1.635 , 0.023]
squared fraction held by outside blockholders	0.444	[-0.240 , 1.128]
Cleaned Data		
log (book value of assets)	-0.192	[-0.278 , -0.105]
log (age)	-0.448	[-0.614 , -0.283]
delaware dummy	-0.014	[-0.242 , 0.215]
SP500 inclusion dummy	1.119	[0.771 , 1.468]
fraction held by outside blockholders	-2.688	[-4.350 , -1.026]
squared fraction held by outside blockholders	3.286	[-0.223 , 6.794]

The table above displays the bootstrapped coefficients and confidence intervals for the 1998 Q regression, as displayed in table 6. The bootstrap method uses the observed sample with $n = 1, \dots, 1,216$ independent observations of our dependent variable ($\mathbf{y} = \mathbf{y}_1, \dots, \mathbf{y}_{1216}$), the industry adjusted Q, and the independent variables ($\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_{1216}$). A paired bootstrap sample is obtained by independently drawing **25,000** pairs $(\mathbf{x}_i, \mathbf{y}_i)$ from the observed sample with replacement. We then estimate a Q regression with each of the 25,000 bootstrapped samples. The resulting slopes are used to construct the confidence intervals and coefficients.

Table 8**1998 Q Regressions : A Comparison of Alternative Fixes of the Raw Data**

Regression Type	CompactD raw data	CompactD raw data	Cleaned Data	Cleaned Data	Winsorized at 50%	Winsorized at 100%	Truncated at 50%	Truncated at 100%	Cleaned at 50% ^d	Cleaned at 100%
	OLS	Median	OLS	Median	OLS	OLS	OLS	OLS	OLS	OLS
Number of observations	1,216	1,216	1,216	1,216	1,216	1,216	1,137	1,194	1,216	1,216
% holding at which data was cleaned, truncated or winsorized	---	---	---	---	50%	100%	50%	100%	50%	100%
% held through outside blocks	-0.81 (0.46)	-0.135 (0.154)	-2.69** (0.91)	-0.97** (0.35)	-1.38 (1.13)	-1.33 (0.69)	-0.80 (1.26)	-1.67* (0.81)	-1.71* (0.87)	-1.66* (0.80)
% held through outside blocks squared	0.44 (0.34)	0.06 (0.08)	3.29* (1.57)	1.01 (0.60)	1.52 (2.37)	1.29 (0.90)	-0.38 (3.04)	2.01 (1.27)	2.21 (1.57)	1.94 (1.26)

This table presents the coefficients of the sum of the fraction of the firm held as outside blocks and squared outside blockholdings, before and after our cleaning procedure, and from various alternative regression setups involving the 1998 cross-sectional regressions of industry-adjusted Tobin's Q on the blockholdings and all of the control variables. Standard errors are reported in parentheses and significance at the one-percent and five-percent levels is indicated by ** and *, respectively.

Figure 1

Compact Disclosure's October 1999 Director and Large Shareholder Data for Coca Cola Co.

```

CO- COCA COLA CO|
TS- KO|
CU- 191216100|

NA- (SOURCE: 10K)^
  IVESTER, M. DOUGLAS/ 51/  CHAIRMAN OF THE BOARD, CHIEF EXECUTIVE OFFICER,
  EXECUTIVE VICE PRESIDENT (PRX 03-04-99) / $12,872,587^
  CHESTNUT, JAMES E./ 48/  GENERAL OFFICER (PRX 03-04-99) / NA^
  STAHL, JACK L./ 46/  GENERAL OFFICER (PRX 03-04-99) / NA^
  DAFT, DOUGLAS N./ 56/  GENERAL OFFICER (PRX 03-04-99) / NA^
  WARE, CARL/ 55/  SENIOR VICE PRESIDENT, GROUP PRESIDENT / NA^
  HAAS, TIMOTHY J./ 52/  SENIOR VICE PRESIDENT, GROUP PRESIDENT / NA^
  COOPER, RALPH H./ 59/  GENERAL OFFICER (PRX 03-04-99) / NA^
  CASEY, WILLIAM P./ 58/  SENIOR VICE PRESIDENT, GROUP PRESIDENT / NA^
  GLADDEN, JOSEPH R., JR./ 56/  SENIOR VICE PRESIDENT, LEGAL COUNSEL / NA^
  FRENETTE, CHARLES S./ 46/  SENIOR VICE PRESIDENT / NA^
  AMON, ANTON/ 55/  SENIOR VICE PRESIDENT, MANAGER / NA^
  GOURLAY, GEORGE/ 57/  SENIOR VICE PRESIDENT, MANAGER / NA^
  WALTERS, MICHAEL W./ 52/  VICE PRESIDENT / NA^
  SHAW, SUSAN E./ NA/  SECRETARY (PRX 03-04-99) / NA|

DO- (SOURCE: PROXY 03/04/1999)^
  IVESTER, M. DOUGLAS/ 51/  CHAIRMAN OF THE BOARD, CHIEF EXECUTIVE OFFICER,
  EXECUTIVE VICE PRESIDENT / $12,872,587^
  ALLEN, HERBERT A./ 59/  DIRECTOR / NA^
  ALLEN, RONALD W./ 57/  DIRECTOR / NA^
  BLACK, CATHLEEN P./ 54/  NA/ NA^
  BUFFETT, WARREN E./ 68/  NA/ NA^
  KING, SUSAN B./ 58/  NA/ NA^
  MCHENRY, DONALD F./ 62/  DIRECTOR / NA^
  NUNN, SAM/ 60/  DIRECTOR / NA^
  OREFFICE, PAUL F./ 71/  DIRECTOR / NA^
  ROBINSON, JAMES DIXON, III/ 63/  DIRECTOR / NA^
  UEBERROTH, PETER V./ 61/  DIRECTOR / NA^
  WILLIAMS, JAMES B./ 65/  DIRECTOR / NA|

SH-
  BUFFETT, WARREN E.,  8.10% (PRX 03-04-99)^
  SUNTRUST BANKS INC,  6.10% (PRX 03-04-99)^
  BERKSHIRE HATHAWAY INC,  8.10% (PRX 03-04-99)^

  ****^
  TYPE          DATE(Q,M)  OWNERS  CHANGE (000S)  HELD  %OWN^
  INSTITUTIONS  03/31/1999(Q)  1      NIL             NIL  0.00^
  5% OWNERS     05/31/1999(M)  0      NA              0   0.00^
  INSIDERS      05/31/1999(M)  0      NA              0   0.00 ||
  
```

This figure reproduces the results of a string search of the October 1999 *Compact Disclosure* CD ROM. The database was searched based on Ticker Symbol, and the fields CO – company name, TS – Ticker Symbol, CU – Cusip number, NA – Officers, DO – Directors, and SH – large shareholders were requested in the tagged format. Below the symbol “****^” at the bottom, the *Spectrum* data are listed automatically by *Compact Disclosure*.

Figure 2

Beneficial Ownership Table from the March 4th, 1999 Proxy of Coca Cola Co.

The following table sets forth information regarding beneficial ownership of Company Common Stock by each Director, the Company's five most highly compensated executive officers and the Directors and executive officers of the Company as a group, all as of February 22, 1999.

NAME -----	AGGREGATE NUMBER OF SHARES BENEFICIALLY OWNED -----	PERCENT OF OUTSTANDING SHARES/18/ -----
Herbert A. Allen.....	9,368,875/1/	*
Ronald W. Allen.....	13,194/2/	*
Cathleen P. Black	15,541/3/	*
Warren E. Buffett.....	200,004,396/4/	8.1%
Susan B. King.....	12,915/5/	*
Donald F. McHenry.....	29,023/6/	*
Sam Nunn.....	2,039/7/	*
Paul F. Oreffice.....	109,252/8/	*
James D. Robinson III.....	16,298/9/	*
Peter V. Ueberroth.....	88,605/10/	*
James B. Williams.....	106,053,931/11/	4.3%
M. Douglas Ivester.....	5,334,915/12/	*
Jack L. Stahl.....	1,282,326/13/	*
Douglas N. Daft.....	969,748/14/	*
James E. Chestnut.....	437,379/15/	*
Ralph H. Cooper.....	1,031,801/16/	*
All Directors and Executive Officers as a Group (25 Persons)	329,575,298/17/	13.4%

[...]

/4/ Shares owned indirectly through subsidiaries of Berkshire Hathaway Inc., the capital stock of which is owned 31.5% by Mr. Buffett and three trusts of which he is trustee but in which he has no beneficial interest and 2.4% by his wife. Also includes 4,396 phantom shares accrued under the Deferred Compensation Plan for Non-Employee Directors.

[...]

SECTION 16(A) BENEFICIAL OWNERSHIP REPORTING COMPLIANCE

PRINCIPAL SHARE OWNERS

Set forth in the table below is information as of December 31, 1998 with respect to persons known to the Company to be the beneficial owners of more than five percent of the Company's issued and outstanding stock:

NAME AND ADDRESS -----	NUMBER OF SHARES BENEFICIALLY OWNED -----	PERCENT OF CLASS -----
Berkshire Hathaway Inc./1/ 1440 Kiewit Plaza Omaha, Nebraska 68131	200,000,000	8.1%
SunTrust Banks, Inc./2/ 303 Peachtree Street Atlanta, Georgia 30308	151,135,261	6.1%

/1/ Berkshire Hathaway Inc., a diversified holding company, has informed the Company that certain of its subsidiaries hold an aggregate of 200,000,000 shares of Company Common Stock. The capital stock of Berkshire Hathaway Inc. is beneficially owned 31.5% by Warren E. Buffett and three trusts of which he is a trustee but in which he has no beneficial interest and 2.4% by his wife. All of such shares of the Company are included in the share ownership of Mr. Buffett disclosed in the table of beneficial ownership of securities above.

/2/ [...]

Figure 3

Compact Disclosure's 1998 Large Shareholder Data for Outback Steakhouse, Inc.

```

CO- OUTBACK STEAKHOUSE INC|
TS- OSSI|
CU- 689899102|

DO- (SOURCE: PROXY 03/13/98)^
SULLIVAN, CHRIS T./ 50/ CHAIRMAN OF THE BOARD, CHIEF EXECUTIVE OFFICER / $400,000^
BASHAM, ROBERT D./ 50/ PRESIDENT, CHIEF OPERATING OFFICER / $400,000^
MERRITT, ROBERT S./ 46/ CHIEF FINANCIAL OFFICER, SENIOR VICE PRESIDENT, TREASURER /
$270,000^
BRABSON, JOHN A., JR./ 57/ DIRECTOR (10-Q 06-30-98)/ NA^
BRIDGES, CHARLES H./ 67/ DIRECTOR (10-Q 06-30-98)/ NA^
GANNON, J. TIMOTHY/ 49/ SENIOR VICE PRESIDENT, DIRECTOR (10-Q 06-30-98)/$290,000^
SELMON, LEE ROY/ 43/ DIRECTOR (10-Q 06-30-98)/ NA^
CAREY, W. R., JR./ 50/ DIRECTOR / NA^
FIELDS-ROSE, DEBBI/ 41/ DIRECTOR / NA^
FLOM, EDWARD L./ 68/ DIRECTOR / NA^
SCHNEID, NANCY/ 39/ DIRECTOR / NA^
WILT, TOBY S./ 53/ DIRECTOR / NA^
AVERY, PAUL E./ 38/ SENIOR VICE PRESIDENT, DIRECTOR (10-Q 06-30-98)/$579,213^
FIELDS, DEBBI/ 41/ DIRECTOR (PRX 03-13-98) / NA^
GANNON, TIMOTHY J./ 49/ SENIOR VICE PRESIDENT (PRX 03-13-98) / NA|

SH-
BASHAM, ROBERT D., 17.09% (PRX 03-13-98)^
SULLIVAN, CHRIS T., 17.39% (PRX 03-13-98)^
GANNON, TIMOTHY J., 17.12% (PRX 03-13-98)^
MULTI VENTURE PARTNERS LTD, 16.85% (PRX 03-13-98)^
T ROWE PRICE ASSOCIATES INC, 9.08% (PRX 03-13-98)^

    ***^
    TYPE          DATE(Q,M) OWNERS  CHANGE (000S) HELD  %OWN^
    INSTITUTIONS  06/30/98(Q)  152      1,406      31,204 63.68^
    5% OWNERS     09/30/98(M)   5         NA       30,409 62.06^
    INSIDERS      06/30/98(M)   9         NA       8,741 17.84 ||
  
```

This figure reproduces the results of a string search of the September 1998 *Compact Disclosure* CD ROM. The database was searched based on Ticker Symbol, and the fields CO – company name, TS – Ticker Symbol, DO – Directors, CU – Cusip number, and SH – large shareholders were requested in the tagged format. Below the symbol “***^” at the bottom, the *Spectrum* data are listed automatically by *Compact Disclosure*.

Figure 4

Beneficial Ownership Table from March 13th, 1998 Proxy of Outback Steakhouse, Inc.

SECURITY OWNERSHIP OF CERTAIN BENEFICIAL OWNERS AND MANAGEMENT

The following table sets forth certain information regarding the beneficial ownership of the Company's Common Stock as of February 27, 1998 (except as noted) by each person known to the Company to own beneficially more than five percent of the Company's Common Stock, each director, each nominee for election as a director, each executive officer, and all executive officers and directors as a group.

Name of Beneficial Owner	Amount Beneficially Owned	Percent of Class	
Chris T. Sullivan	8,505,490	17.39%	1)
Robert D. Basham	8,360,875	17.09%	2)
J. Timothy Gannon	8,373,761	17.12%	3)
[...]			
Multi-Venture Partners, Ltd.	8,242,125	16.85%	0)
T. Rowe Price Associates, Inc.	4,442,400	9.08%	4)

[...]

0) Multi-Venture Partners, Ltd. ("MVP") is an investment partnership formed by Chris T. Sullivan, Robert D. Basham and J. Timothy Gannon. Messrs. Sullivan, Basham and Gannon are the only limited partners in MVP and are the only members of MVP's sole general partner, SBG Investments, L.L.C. ("SBG"), a limited liability company.

[...]

1) Includes (i) 8,242,125 shares owned by MVP; (ii) 231,292 shares owned by Sullivan Family Investments, Ltd., a family limited partnership of which Mr. Sullivan serves as general partner; and (iii) 1,712 shares owned by Mr. Sullivan's children for whom Mr. Sullivan serves as custodian. Mr. Sullivan shares voting and dispositive power with respect to Common Stock owned by MVP

2) Includes 8,242,125 shares owned by MVP. Mr. Basham shares voting and dispositive power with respect to Common Stock owned by MVP.

3) Includes 8,242,125 shares owned by MVP. Mr. Gannon shares voting and dispositive power with respect to Common Stock owned by MVP.

[...]

4) Based on a Schedule 13G filed by T. Rowe Price Associates, Inc. ("T. Rowe Price") with the Securities and Exchange Commission on February 10, 1998, these securities are owned by various individual and institutional investors for which T. Rowe Price serves as investment adviser with power to direct investments and/or sole power to vote the securities. For purposes of the reporting requirements of the Securities Exchange Act of 1934, T. Rowe Price is deemed to be a beneficial owner of such securities; however, T. Rowe Price expressly disclaims that it is, in fact, the beneficial owner of such securities.

Figure 5

Compact Disclosure's 2000 Large Shareholder Data for Tribune Co.

CO- TRIBUNE CO|
 TS- TRB|
 CU- 896047107|

SH-

NORTHERN TRUST CO, 9.10% (PRX 03-27-01)^
 NORTHERN TRUST CO, 100.00% (PRX 03-27-01)^
 ROBERT R MCCORMICK TRIBUNE FOUNDATION, E, 14.95% (PRX 03-27-01) ^
 CHANDLER TRUST NO 1, 12.29% (PRX 03-27-01)^

Figure 6

Beneficial Ownership Table from the March, 27th, 2000 Proxy of Tribune Co.

Principal Shareholders

The following table sets forth information as of February 28, 2001 with respect to each person who is known to Tribune management to be the beneficial owner of more than 5% of any class of Tribune stock entitled to vote:

NAME AND ADDRESS OF OWNER	COMMON STOCK		SERIES B PREFERRED STOCK	
	NUMBER OF SHARES	PERCENT OF CLASS	NUMBER OF SHARES	PERCENT OF CLASS
Robert R. McCormick Tribune Foundation Cantigny Foundation(1) Room 770 435 North Michigan Avenue Chicago, IL 60611	44,825,676	14.95%	--	--
The Chandler Trusts(2) 350 West Colorado Boulevard Suite 230 Pasadena, CA 91105	36,861,190	12.29	--	--
The Northern Trust Company(3) 50 South LaSalle Street Chicago, IL 60675	28,962,648 (4)	9.10%	1,150,456	100%

(1) The investment and voting power of each of the Robert R. McCormick Tribune Foundation and the Cantigny Foundation is vested in a board of six directors, consisting of Dennis J. FitzSimons, Jack Fuller, John W. Madigan and three former Tribune officers.

(2) [...]

(3) On February 28, 2001, The Northern Trust Company, as ESOP trustee, held 2,046,670 shares of Tribune common stock on behalf of the ESOP and was deemed to hold 18,407,296 shares of Tribune common stock into which the Tribune Series B preferred stock is convertible, which shares are included in determining the percent of class owned. All ownership attributed to Northern Trust in its capacity as ESOP Trustee is shared with the participants in the ESOP.

(4) Holdings based upon information contained in a Schedule 13G filed with the Securities and Exchange Commission on February 8, 2001 by Northern Trust, which indicated that Northern Trust had sole voting power with respect to 6,325,093 shares; shared voting power with respect to 22,553,329 shares; sole dispositive power with respect to 6,029,468 shares; and shared dispositive power with respect to 22,564,887 shares.

Figure 7

Empirical distribution of bootstrapped slopes of outside blockholdings in the 1998 Q regression before and after cleaning (N=25,000)

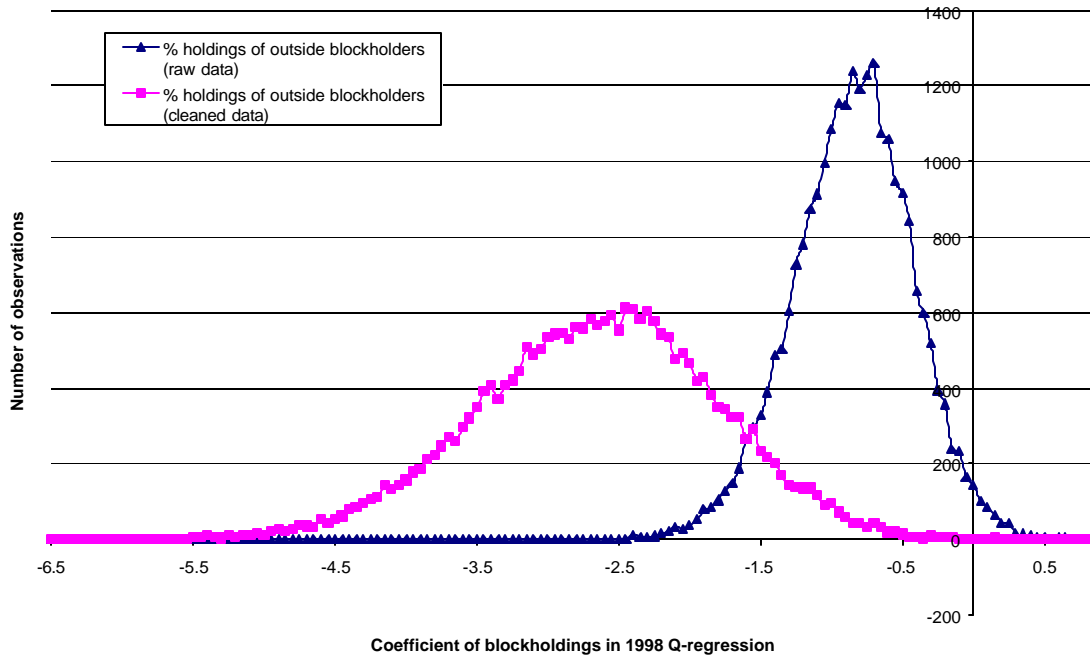


Figure 8

Empirical distribution of the bootstrapped slopes of squared outside blockholdings in the 1998 Q regression (N=25,000)

