Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Corporate Finance 12 (2006) 594–618
Journal of CORPORATE FINANCE
www.elsevier.com/locate/jcorpfin

# Large blocks of stock: Prevalence, size, and measurement

Jennifer Dlugosz [a], Rüdiger Fahlenbrach [b], Paul Gompers [c], Andrew Metrick [d,*]

[a] *Department of Economics and Harvard Business School, Harvard University, USA*
[b] *Department of Finance, Fisher College of Business, Ohio State University, USA*
[c] *Harvard Business School, Harvard University and NBER, USA*
[d] *Department of Finance, Wharton, University of Pennsylvania and NBER, USA*

Received 16 July 2004; accepted 22 April 2005
Available online 27 July 2005

**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 Disclosur*e, 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%, 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

* Corresponding author.
*E-mail address:* metrick@wharton.upenn.edu (A. Metrick).

0929-1199/$ - see front matter 
doi:10.1016/j.jcorpfin.2005.04.002

truncating or winsorizing the sample can reduce about half of the bias in our representative application.


*JEL classification:* G3, Corporate finance and governance
*Keywords:* Ownership; Corporate governance; Blockholders

## 1. 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, IPO underpricing, organizational efficiency, 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 Mørck 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.

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

[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), Field and Sheehan (2004), Denis and Denis (1994), Cremers and Nair (in press), Bhagat et al. (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.