LARGE SHAREHOLDER DIVERSIFICATION AND CORPORATE CASH HOLDING

Roberto Mura[a]

December 2023

ABSTRACT

Using data for almost 400,000 European firms covered in *Orbis*, I reconstruct the portfolios of shareholders who hold equity stakes in private and publicly traded firms between 2007 and 2020. I document a novel relation between the diversification of large shareholders and firms' cash holdings. Firms controlled by more diversified large shareholders tend to hold significantly less cash than those controlled by less diversified shareholders. The impact of large shareholder diversification on cash holdings is economically and statistically significant. The results are robust to controlling for the more conventional proxy of *segment diversification* using a variety of econometric techniques.

JEL Classifications: G11, G15, G31

Keywords: Cash holding; Large shareholders; Portfolio diversification

[a] Alliance Manchester Business School, University of Manchester.
Acknowledgments: I thank Evgeny Lyandres, Maria-Teresa Marchica, and Roni Michaely for insightful comments on a previous version of this paper and Olga Kolokolova for insightful comments on the current version. I am also grateful for the feedback received from seminar participants at the International Finance and Banking Society 2023 and the International Corporate Governance Society 2023.

1. Introduction

It is widely documented that more diversified firms tend to hold less cash. In a highly influential paper, Duchin (2010) shows that companies that operate in multi-segments (diversified firms) hoard less cash than undiversified ones. Similarly, Subramaniam et al., (2011) also report that diversified firms hold significantly less cash than their focused counterparts whereas Tong (2011) shows that the value of cash is lower in diversified firms than in single-segment firms. More recently, using a dynamic investment model, Bakke and Gu (2017) confirm this general result.

A common thread of this literature is the focus on *segment* diversification at the *firm* level. In my study, I offer a different and novel perspective by focusing on diversification at the *shareholder* level rather than at the *firm* level. I argue that a dominant large shareholder with a less diversified portfolio will be more concerned with the firm's cash flow variability. Higher cash holdings at the firm level would reduce her exposure to this risk. Consequently, ceteris paribus, companies controlled by less-diversified shareholders are expected to accumulate more cash than those controlled by more-diversified shareholders.

Several studies in recent years have offered compelling evidence that diversification at the shareholder level is an important determinant of firm choices on several grounds. For instance, Bodnaruk et al., (2008) provide evidence that firms held by less diversified controlling shareholders are more likely to go public and exhibit higher underpricing. Faccio et al., (2011) document that firms controlled by diversified large shareholders undertake riskier investments than firms controlled by non-diversified large shareholders. More

recently, Lyandres et al., (2019) report a very strong and significant association between large shareholders diversification and investment levels.

My study covers almost 400,000 companies across 41 European countries using data from the Bureau Van Dijk's *Orbis Historical* database for the 14-year interval between 2007 and 2020. In my sample, the largest (ultimate) shareholder has an average of 61% of voting rights in their firms. Therefore, it is safe to assume that she has control over the firm's decisions.

I adopt several techniques to test the association between portfolio diversification at the shareholder level and cash holding at the firm level. I follow Faccio et al. (2011) and Lyandres et al. (2019) and use two proxies to measure large shareholder diversification: (i) the (natural log of the) number of firms in which the controlling investor holds shares across all countries in the sample and (ii) the Herfindahl index of wealth concentration at the investor level.

I find strong evidence that higher diversification at shareholder level is associated with lower levels of cash holding. The association is not only statistically significant but also *economically large.* Using the investor fixed effect specification as a baseline model, one standard deviation increase in the level of portfolio diversification (measured by the natural logarithm of the number of firms held) results in a 22% decrease in cash holding relative to its mean. The results are robust when I use the alternative proxy for portfolio diversification (measured by the Herfindahl index of wealth concentration). Moreover, results are robust to the inclusion of a proxy of firm level segment diversification. I find that, while still statistically significant, the economic effect of firm-level segment diversification is smaller than that of portfolio diversification of the large shareholder.

A potential concern with my empirical design could be that results may be driven by endogeneity and in particular self-selection. In other words, it could be argued that shareholders *select* to invest in firms with a level of cash holding which best suits their preferences – like risk tolerance for instance – rather than influencing the cash decisions of these firms. In my sample, such a selection mechanism is, however, highly unlikely. Almost 100% of the sample is made of private and illiquid companies in which, as discussed earlier, the average cash flow and voting rights are about 60% (with medians of 51%). It is very hard to imagine that these large shareholders would frequently and proactively adjust their portfolios. Nonetheless, I take several steps to address the potential endogeneity concerns.

First, I show that the positive association between shareholder portfolio diversification and cash holding is robust in a panel regression framework, in which I include shareholder fixed effects, as well as the more conventional country, industry and year-fixed effects. The inclusion of shareholder fixed effects has the benefit of controlling for investor-specific (time-invariant) omitted variables that may affect the investor's attitude to risk, which may drive the decision to invest in a more or less cash-rich company. I also perform a separate set of tests controlling for firm fixed effects, that is, *time-invariant* firm-specific characteristics that may be correlated with omitted explanatory variables.

Second, to further stress-test whether my results are driven by endogeneity, I employ an instrumental variable technique. In the first instance, I follow the papers of Faccio et al., (2011) and Laeven and Levine (2007) and Laeven and Levine (2009) and use the average portfolio diversification of large shareholders of all the other companies in the same country, year and industry as an instrumental variable for each shareholder's degree of portfolio diversification. Given the nature of this IV, there should be no association between the cash

holding of a company with the diversification of other companies' shareholders and therefore this should satisfy the exclusion restriction.

To further minimize possible concerns of endogeneity, I repeat the analysis by mechanically breaking the link between the instrument and the firm at a country level. I perform a matching exercise to find for each company in the sample a similar company in a different country in the same year. Then, to instrument portfolio diversification of a given shareholder, I use the average portfolio diversification of all other shareholders in the country and year of a matched firm. Thus, for each firm in the sample, the instrument comes from the matched firm in a different country. In this setup, any association between the dependent variable and the instrument is very unlikely.

As a further step, I also employ a variation of the Heckman two-step approach: the treatment effects model. My choice of an exogenous determinant of the propensity to diversify the portfolio is motivated by the findings in Lyandres et al., (2019), and I use the average number of companies that are located within a certain geographical distance from the large shareholder as an instrument for the diversification of the portfolio of the largest shareholder. This proxy should represent a set of potential "investable" firms for my sample of (mostly privately held illiquid) companies. Portfolio diversification of the largest shareholder should be positively related to the number of investable firms. At the same time, this measure based on geographical distance should not be related to the level of cash holding at the firm level and therefore it should also satisfy the exclusion restriction.

In the last step, I perform a very stringent matching exercise to isolate matched samples of statistically indistinguishable diversified and non-diversified firms, in the attempt to minimize further possible concerns of endogeneity. Results clearly indicate that, everything else (controllable) being

equal, diversified firms tend to hold less cash than non-diversified ones. While taken individually none of these steps perfectly addresses endogeneity, they all confirm my main conclusion.

My paper contributes to several strands of literature. First, I offer a different perspective to the existing evidence on the impact of diversification on cash holding. I shift the focus from *firm* diversification at the sector level, explored in several previous studies, to the diversification of the portfolio at the *shareholder* level. My results show that the higher the degree of diversification of the portfolio of the largest owner, the lower the average level of cash holding at the firm level. All tests include a more conventional proxy for *firm* diversification at the sector level, and the results show that both mechanisms are present and relevant to firms' financial policy decisions.

Second, my paper contributes to the strand of literature that documents the links between shareholder diversification and firm decisions. John et al., (2008) find no significant relation between ownership concentration and corporate risk-taking. Bodnaruk et al., (2008) provide compelling evidence that firms held by less diversified controlling shareholders are more likely to go public, and they exhibit a higher level of underpricing. Faccio et al., (2011) show how firms controlled by diversified large shareholders undertake riskier investments than firms controlled by non-diversified large shareholders. Lyandres et al., (2019) provide evidence that owners' portfolio diversification is a strong predictor of the level of investment. My study unveils a further link between shareholder diversification and firm decisions: the cash holding policy.

Third, my paper extends the growing literature that focuses on privately held companies. Several studies have documented how private and public firms differ in many ways, from their financing decisions to their ownership and governance structures. Mortal

and Reisel (2013) document that listed firms are more capable to pursue positive NPV projects than private firms. Asker et al., (2015) on the other hand report that public firms invest generally less, and they are also less responsive to changes in investment opportunities possibly due their higher sensitivity to short-term pressures. More recently, Lyandres et al., (2019) report that European public firms invest more than private ones, similarly to Mortal and Reisel (2013).

With emphasis on cash holding, Gao et al., (2013) report that in the US, private companies tend to keep around half the cash compared to a *matching* sample of public companies. Similarly, Mortal et al., (2020) find that European private firms hold less cash than a *matching* sample of public firms and relate this difference to different borrowing costs. On the contrary, Hall et al., (2014) report that European private firms (mostly from emerging markets) tend to display higher cash balances than an *unmatched* sample of publicly listed companies. While the focus of my study is not on private vs public firms, my results consistently indicate that all else being equal, private firms hold less cash than public firms.

The rest of the paper is organized as follows. Section 2 presents the data and descriptive statistics. Section 3 presents regression results and addresses possible endogeneity concerns. Section 4 reports a battery of robustness tests, and Section 5 summarizes the findings and concludes.

2. Data and variables

2.1 Sample

I collect direct ownership and accounting data for all companies included in the *Orbis Historical* database from Bureau Van Dijk. Accounting data are available from 1999, while ownership information is only available from 2007. After determining the largest ultimate owner for each firm (more on this below), I apply two main filters to the data.

First, I keep only countries that are present in Amadeus. The reason for applying this filter is two-fold. First, as discussed above, these data have been used in recent studies, and their quality has been confirmed (see Faccio et al., 2011). Second, this filter is applied for computational reasons. Back-of-the-envelope calculations on the full Orbis sample show that after filtering out financial firms and firms with missing cash flow data, there are about 30,245,157 non-financial firms for 153 countries for a total of 206,498,760 observations.

Second, I apply a size filter: for all countries, the average total assets in the time series of the firm must be at least 5 million euros. This is done to reduce the dominance of companies with only one shareholder in smaller firms. After applying these filters, the final sample consists of 393,691 companies and 2,889,134 observations for 41 countries, spanning between 2007 and 2020.[1]

[1] Belarus, Liechtenstein and Monaco are the only countries in the original Amadeus list that disappear due to the filters imposed on the data.

2.2 Measures of portfolio diversification

Direct ownership information is collected from the ownership section of the Orbis Historical. For each company with available ownership data, I identify all ultimate shareholders. That is, whenever the direct shareholder of a firm is another company, I identify its owners, the owners of its owners, and so on. Once the full chain of control is identified, I define the voting rights of the ultimate owner as the weakest link along the chain of control. The cash flow rights are defined as the product of all the links in the chain of control. This approach is consistent with earlier studies like Claessens et al., (2000), Faccio and Lang (2002), Faccio et al., (2011) and Lyandres et al., (2019). An improvement over prior studies is that there is no limit on firm size before the ultimate owner is calculated. Therefore, I am able to measure smaller ownership stakes than previous studies.[2] The filter on the size of the firms is only applied at the end.[3]

I also calculate the spread between voting rights and cash flow rights. I use the spread to address the possibility that some of my results may reflect tunneling because a high level of spread may give incentives to the controlling shareholders to expropriate minority shareholders.

After tracing all ultimate shareholders and their cash flow and voting rights, I identify the shareholder controlling the largest fraction of voting rights in each firm. The ownership,

[2] Faccio et al., (2011) and Lyandres et al., (2019) use Amadeus top 250,000 which did not allow access to data on smaller firms.

[3] I apply no filter before calculating ultimate control, ownership or diversification proxies. However, I then remove from the final sample companies where the largest owner is the State, Government or any kind of public authority. I also remove firms where the owner is a Foundation, Financial Institution (like Banks, Insurance Companies, Hedge Funds or Mutual Funds) or if they are reported as “unclassified”.

control, and diversification variables employed throughout the paper always refer to each firm's largest ultimate shareholder since these would be the most likely candidates to drive the decisions and policies at the firm level given their voting power.

The first measure of portfolio diversification, *Ln No. Firms*, is the natural log of the number of companies in which a company's largest ultimate shareholder holds shares, directly or indirectly, each year. The second proxy is a measure of wealth concentration: the *Herfindahl* index. To compute the Herfindahl index, I first collect the book value of the equity corresponding to each equity position in the portfolio of the ultimate owner. These are multiplied by the corresponding values of cash flow rights. I then compute the weight of each stock in the portfolio. The Herfindahl index is the sum of squared weights. The index ranges from 0 to 1, with 1 indicating that all wealth is invested in one firm (fully concentrated wealth), and 0 indicating a totally diversified portfolio. In the analysis, I use the transformation (1-Herfindahl) index, so that higher values represent higher diversification in line with the first proxy.

2.3 Economic variables

The dependent variable in my study is the classic proxy for cash holding defined as the ratio of cash and cash equivalent to total assets. I add several controls to the cash holding model to minimize the risk of omitted variable bias. In line with previous studies, I include the following determinants of cash holdings: (1) *Private Company*, a dummy variable taking the value of one when the firm is privately held and zero when the firm is listed on a stock exchange; (2) *Growth Opportunities*, defined as annual growth in total assets, where total

assets is the sum of fixed assets (tangible and intangible fixed assets and other fixed assets) and current assets (inventory, receivables, and other current assets); (3) *Cash Flow*, defined as the ratio of income plus depreciation to total assets; (4) *Cash Flow Volatility*, defined as the standard deviation of the cash flows at the country-industry level over five years overlapping; and (5) *Ln (1+Age)*, defined as the natural log of (1 + the number of years since incorporation). This variable controls for differences in the life cycle of firms, as one would expect that younger firms may face stronger financing frictions and, hence, hold more cash ; (6) *Ln(TA)* is defined as the natural log of total assets; (7) *Leverage* is defined as the ratio of total debt to total assets, where total debt includes noncurrent liabilities (long-term debt and other noncurrent liabilities) and current liabilities (loans, accounts payable, and others); (8) *Spread* measures the difference between voting rights and cash flow rights of the ultimate owner because a high divergence entails stronger expropriation incentives for the ultimate owner; and (9) *Ln*(*Sector Diversification)* is the natural log of the number of business segments reported by the firm.

All data are winsorized to minimize the impact of outliers. All financial firms are removed from the data and all observations where any of the controls are missing are also discarded.

2.4 Descriptive statistics

Table I reports detailed descriptive statistics. Panel A reports the statistics of the ownership variables at the investor level (to avoid possible duplications) while Panel B reports the rest of the controls at the firm level. The mean number of firms in the portfolio of

the largest owner is 9.171 while the median is 3. The variable exhibits considerable variation since the standard deviation is about 42. In the top 25% of the distribution, I find owners with seven or more equities in their portfolio; in the top 10%, I find owners with 15 or more equities; in the top 5%, they have 27 or more equities, and the top 1% have 102 or more equities.

The mean and median cash flow and voting rights are in the region of 60% (51%), while the spread (the difference between Voting Rights and Cash Flow Rights) is essentially zero. Only in the top 5% of the distribution do we find firms with a spread of approximately 10%. This indicates that these largest shareholders are dominant owners with no incentive to expropriate the firm(s) they control. These statistics are relatively close to those reported in earlier studies such as Lyandres et al. (2019) and Faccio et al. (2011), considering the availability of smaller firms in this study.

Panel B reports descriptive statistics for the controls at the firm level. The sample firms hold an average of 10% of their assets in the form of cash holdings. Given the nature of the sample, no market value is available, since about 99% of the sample comprises private firms. Therefore, Growth Opportunities are approximated via the growth in Total Assets, which is approximately 16% for an average firm in the sample.[4] Leverage is about 64% of the total assets, while the average firm in the sample is 23 years old.

[4] Results are robust to using growth in sales although this variable contains more missing values which is why I opted for growth in total assets. I replicate the main test using this alternative control in the Appendix.

The proxy for diversification at the industry level also displays a meaningful level of variability. While the average is 1.7, and the median is 1, the top 25% of companies operate in two or more industries, the top 10 % in three or more, while the top 1% operate in four or more.

Figure 1 plots the time series of the average levels of cash holdings for firms in which the largest investor only has equity in one company (Non_Diversified) versus those companies in which the shareholder has equity positions in two or more companies (Diversified). The two series follow parallel increasing trends, although the difference between the two cohorts increases over time. In the early years, the difference was about 5%, while in the latest year it peaked at about 9%, with Non-Diversified firms having lower levels of cash.

3. Regression Analysis

To investigate the association between the largest shareholder's portfolio diversification and corporate cash holding, I present three main sets of tests that differ in the types of fixed effects included in the model. All models are estimated using robust standard errors.

My first regression equation is:

$$Cash_{i,t} = \alpha + \beta\ Diversification\ i_{,t} + \gamma X_{i,t} + Industry\ F.E. + Country\ F.E. + Year\ F.E. + u_{i,t} \quad (1)$$

where *Diversification* is one of the two measures described above; $X_{i,t}$ is a vector of control variables, which, as discussed earlier, includes *Private Company*, *Growth Opportunities*, *Cash Flow*, *Cash Flow Volatility*, *Ln(1+Age)*, *Ln(TA)*, *Leverage*, *Spread*, *and Ln(Sector Diversification).* Model 1 above also includes year fixed effects (always included across all specifications presented in the paper) to control for the effect of possible macro-trends, Industry Fixed Effects (SIC4) and Country Fixed Effects.

The second set of regression tests further exploits the panel dimension of the dataset at the shareholder-level. The panel regressions allow me to control for unobservable shareholder-specific characteristics that may potentially impact a firm's cash holdings by including shareholder-fixed effects. For example, it is possible that different shareholders may have different levels of risk tolerance not fully captured by my diversification proxies. More generally, the inclusion of shareholder fixed effects allows me to control for any shareholder-specific characteristic that may be correlated with the omitted explanatory variables. Controlling for shareholder fixed effects then helps reduce possible concerns of omitted variable bias. In the second set of tests, the regression equation is:

$$Cash_{i,t} = \alpha + \beta\ Diversification\ i_{,t} + \gamma X_{i,t} + Industry\ F.E. + Country\ F.E. + Shareholder\ F.E. + Year\ F.E. + u_{i,t} \quad (2)$$

In the last specification, I follow standard practice and control for firm fixed effects (rather than shareholder fixed effects). The purpose is the same: to reduce concerns of omitted variable bias, which may taint the validity of the regression tests.

In this third set of tests, the regression equation is:

$$Cash_{i,t} = \alpha + \beta\ Diversificationi_{,t} + \gamma X_{i,t} + Firm\ F.\ E.\ + Year\ F.E.\ + u_{i,t} \quad (3)$$

Table II reports separate results for both portfolio diversification measures. The negative coefficients on *Ln No.Firms* and (*1-Herfindahl*) confirm my mainline hypothesis and the univariate result reported in Figure 1. Companies that are controlled by better-diversified owners are more risk tolerant and therefore are less in need to store cash – all else being equal. The association with cash holdings is not only statistically significant, but also economically large. A standard deviation increase in either proxy of portfolio diversification leads to a reduction in cash holdings by approximately 20% (Models 3 and 4) for the average firm. In my sample, the effect of shareholder diversification on firm cash holdings is similar in magnitude, if not larger, to the economic impact of sector diversification, as documented in previous studies. For instance, back of the envelope calculations from Duchin (2010) suggests that according to his tables, a standard deviation increase in "Number of segments" decreases average cash by about 6.7% of the mean. My findings therefore complement the existing important studies, such as Duchin (2010), Subramaniam et al., (2011), Tong (2011), Fernandes and Gonenc (2016), Bakke and Gu (2017), and Gu (2017), by offering an alternative and original perspective on the determinants of firm cash holdings.

Other variables behave relatively in line with mainstream hypotheses (e.g., Bates et al., 2009). I find that larger firms appear to hold less cash than smaller firms. Possibly, economies of scale allow larger firms to face lower transaction costs, leading them to store lower amounts of liquid funds in relative terms. Leverage is negative and significant,

suggesting that, to some degree, these sources of funds act as substitutes. As discussed in Bates et al., (2009): “If debt is sufficiently constraining, firms will use cash to reduce leverage, resulting in a negative relation between cash holdings and leverage.” This line of argument may be very pertinent in my case since the super-majority of companies are privately held and are expected to face tighter borrowing constraints than listed firms.

3.1 Endogeneity concerns

In the previous section, I tried to address endogeneity concerns arising from omitted variables by controlling for time-varying observables that may affect both cash holdings and portfolio diversification. I added investor fixed-effects to the regression specifications to control for time invariant unobservables that differ across large shareholders. I also added firm fixed-effects to control for time invariant firm specific unobservables that may lead to biased and inconsistent estimates.

Another possible endogeneity concern is related to the direction of causality in the results. A potential feedback effect from the level of firm cash holdings on the portfolio diversification of the largest shareholder would imply reverse causality. For example, investors planning to invest in a more (less) cash-rich firm would therefore adjust the structure of their holdings to increase (decrease) portfolio diversification.

This interpretation of my results implies frequent changes to portfolios held by large shareholders that are not observed in the data. Almost 99% of the firms in the sample are illiquid privately held companies, and the mean/median ownership position in these

companies is 60% (51%). It is very hard to imagine that these large shareholders would adjust their large illiquid equity positions rather than simply adjusting the cash holding of the firms they (fully) control. That being said, I report two formal tests addressing the reverse causality issue, both based on an instrumental variable technique.

3.1.1 Instrumental Variables

In the first instance, I follow Faccio et al., (2011) and Laeven and Levine (2007) and Laeven and Levine (2009) and calculate the average portfolio diversification of large shareholders of all the other companies in the same country, year, and industry. This variable is meant to capture the "natural" tendency to diversify across all large shareholders involved in similar types of activities. At the same time, this variable should not play any direct role in shaping a company's cash position, as it is calculated for all other shareholders (in the same country year and industry), excluding the firm itself.

The results of this test are listed in Table III. For brevity from now on, I report only the results with either investor- or firm-fixed effects. Odd-numbered models represent second-stage regressions, whereas even-numbered models report the first stage. In the first-stage regressions, I include all exogenous variables, along with the instrumental variable, to explain a large shareholder's actual diversification choice. I report the F-statistic and partial R^2 for the instruments in the first-stage regression. As shown in Model (2) of Table III, the "natural" degree of portfolio diversification is positively and strongly related to the endogenous variable, with an F-stat of 1329 and a partial R^2 of 0.363. As a rule of thumb, an

F-statistic below 10 would suggest a weak instrument, as discussed by Staiger and Stock (1997). All models in Table III display similar values, and therefore alleviate possible concerns that my coefficient estimators suffer from biases due to having weak instruments (Bound et al., 1995). More importantly, across all models, the proxies for portfolio diversification of the largest shareholder retain a negative and significant coefficient, confirming the main results reported in Table II.

3.1.2 "Scrambled" Instrumental Variables

To further minimize possible endogeneity concerns, I repeat the analysis by (further) mechanically breaking the link between the instrument and the firm. In the first stage, I perform a simple matching exercise to find a similar company in a different country in the same year for each company in the sample. For each firm in the sample, the instrument is then derived from a matched firm in another country. In this way, I deliberately "mismatch" the cash holding and portfolio diversification variable with the IV.

To make a practical example, let us call IV1 the instrument described in Section 3.1.1: the average portfolio diversification of large shareholders of all other companies in the same country year and industry. In this first IV regression, firm A's cash holding in Country X in year N is regressed on the portfolio diversification of the largest owner of firm A in Country X in year N, instrumented by the average diversification of all the other large shareholders in Country X (in Sector S) in year N.

To compute the “scrambled” IV2, I match Firm A in Country X to (say) Firm B in Country Y in year N. The scrambled IV2 for Firm A in Country X is the actual IV1 of Firm B in Country Y in year N. Given the scrambling between firms and IVs, the data entry for this modified instrument comes from a matching firm in a different country. Mechanically, there can be no association between the dependent variable and the instrument with this setup.

Matching is performed on diversification only for simplicity within the same year. I first sort countries by their number of observations and pair them according to this criterion to have groups with treated and control firms with relatively similar numbers. This method ensures minimal loss of data. Also, to ensure that the firms in the “treated” country are very similar to the firms in the “control” country, I impose a maximum difference between the propensity scores of the treated and control of 0.01.

Table IV presents the results of this alternative test. As in the previous tables, I report both stages of the IV regression with odd-numbered models representing the second-stage regressions, while even-numbered models report the first stage. Inevitably, the sample size drops compared to Tables II and III owing to the extra step of matching between pairs of countries with uneven numbers of observations. That being said, this last table confirms previous results. While the instrument is still positive and significant in the first stage, *Ln No.Firms* and (*1-Herfindal*) both still display negative and statistically significant coefficients, confirming that firms with large shareholders that are more diversified tend to hold less cash, everything else being equal.

3.2 Heckman Two Step

In this section, I endogenize the diversification status of the largest shareholder by estimating the variation in the two-stage Heckman (1979) selection model. In the first stage, I run a probit regression, where I model the diversification status of the shareholder with a binary variable indicating whether the number of equity positions is one or more. In the second stage, I re-estimate the baseline model augmented by the Inverse Mills Ratio from the first-stage regression to correct for potential self-selection. This is similar to, for instance, Campa and Kedia, (2002).

A crucial problem in this type of setting is the inclusion (or not) of valid instrumental variables in the first-stage regressions, which are linked to the selection probability but not to the outcome variable. Li and Prabhala (2007) argue that the inclusion of exogenous instruments is not strictly necessary in a treatment effects model as identification is achieved by non-linearity. Accordingly, I first employ a selection model with no exclusion restrictions. Second, I follow Lyandres et al., (2019) and use the average number of companies that are located within a certain *geographical distance* from the large shareholder as an instrument for the diversification of the portfolio of the large shareholder. This proxy captures the "limit" to diversification. The lower the number of companies within a certain geographical distance, the harder it is for a given shareholder to invest in numerous (mostly private) companies. At the same time, this measure based on geographical distance should not be directly related to the level of cash holdings at the firm level; therefore, it should satisfy the exclusion restriction.

This variable is calculated as follows: for each controlling owner with available address or postal code information, I derive information on latitude and longitude. I repeat

the same procedure for each firm included in the sample. Using the available latitude and longitude information, I compute the spherical distance between each investor in each country in each year for all firms in the same country and year. The spherical distance is calculated as follows:

$$d_{j,i} = arccos(cos(lat_j) \times cos(lon_j) \times cos(lat_i) \times cos(lon_i) + cos(lat_j) \times sin(lon_j \times cos(lat_i) \times sin(lon_i) + sin(lat_j) \times sin(lat_i)) \times r,$$

where lat and lon refer to the latitude and longitude in radians, respectively, and r is the radius of Earth in miles. Results from this set of tests are reported in Table V. As mentioned above, for brevity, I only report models that include either investor or firm fixed effects in the second stage, as these are deemed more robust. In Models (1), (2), (5), and (6), I report the results produced via a first-stage probit where no exclusion restriction is included. In models (3), (4), (7), and (8), the results are produced via a first-stage probit, where the average number of companies located within a 50 miles radius of each large shareholder in each country in each year is used as an exclusion restriction in the first stage.[5]

As discussed earlier, information on geographical location is not as populated as standard accounting information; therefore, the sample size decreases.[6] That said, after correcting for the endogenous decision to diversify the portfolio, both variables of

[5] Results are robust to various thresholds of distance. I report two examples in the Appendix using 5 and 25 miles.

[6] Results on a first stage with no exclusion restriction, which preserves full sample size, are virtually identical with those presented in Table V. This is reported in the Appendix.

diversification still display the same negative and highly significant coefficients reported in previous tables.[7]

3.3 Propensity Score Matching

As a last step to try and reduce possible concerns that endogeneity is driving my results, I perform a stringent matching exercise (Rosenbaum and Rubin, 1983). This methodology allows me to identify a matched sample of firms that have an undiversified shareholder and exhibit no *observable* differences in characteristics relative to firms that have more than one equity position in their portfolio (diversified).

In the first step, I run a logit model to calculate the probability (i.e., propensity score) that a firm with given characteristics has a diversified/undiversified large shareholder. More specifically, the propensity score is estimated within each *discrete* measure to ensure the best possible matching. Therefore, this is performed within each country, industry, year, public/private status, age, and segment diversification. This ensures exact matching within each group. In the logit model, I then add all non-discrete controls: *Growth Opportunities*, *Cash Flow*, *Cash Flow Volatility*, *Ln(TA)*, *Leverage* and *Spread*. To ensure the quality of the matching, I impose a maximum difference in the propensity score (caliper) between the treated and control firms to not exceed 0.001 in absolute value. This ensures the similarity between the two groups. The results are reported in Table VI. The very stringent matching

[7] The inclusion of the exclusion restriction in the first stage produces almost identical results to the tests which do not include it in the first stage. The minor differences in coefficients are not visible with max three decimals as in the table. Allowing for more decimals renders these differences visible.

procedure results in a significantly smaller sample, but this ensures a higher quality of matching. This is evident from the comparison of the p-value of the difference between the propensity to be treated (P-score) of 0.897. Moreover, the p-values of the differences in means across all other controls range from a minimum of 0.58 for leverage to a value of 1 for discrete variables. Importantly, the p-value of the test for the difference in means for cash holdings is zero. This strongly indicates that the average cash for companies with a dominant diversified shareholder (7.38%) is statistically different (lower) from that of companies with a dominant undiversified shareholder (17.40%). This result further corroborates our previous findings.

As discussed earlier, while taken individually none of these steps perfectly addresses endogeneity, they all confirm my main conclusion.

4. Robustness Tests

In this section, I perform a series of robustness tests to reduce possible concerns that some confounding effects may be at play and partly explain my results.

4.1. The Role of Dual-Class Shares

One possible limitation of my study is that I am not able to control for the presence of dual class shares in the sample. These are potentially important, since they could significantly impact the calculation of voting rights and cash flow rights, and therefore, the correct identification of the largest ultimate owner. However, several studies have shown that

the use of dual-class shares has been decreasing. For instance, according to Lauterbach and Pajuste (2015), 'The European Union has debated extensively a potential mandatory one share one vote law.' Even though a law was never imposed, Maury and Pajuste (2011) document that in seven European countries with more prevalent use of dual-class shares (Denmark, Finland, Germany, Italy, Norway, Sweden, and Switzerland), the fraction of dual-class (listed) firms decreased from 43% to 29% between 1996 and 2002. Lauterbach and Pajuste (2015) report that, by 2012, the fraction of dual-class shares declined further to 16%.

To reduce concerns that dual-class shares may affect my tests, Table VII reports results where I exclude the seven countries listed above, and the results are essentially unaltered.

4.2. The Role of Tunnelling

As discussed above, when voting rights are meaningfully larger than cash flow rights, the ultimate owner may have incentives to expropriate firms. However, the descriptive statistics clearly indicate that in the vast majority of cases, the spread between voting rights and cash flow rights in the sample is minimal. Inspecting the data in more detail reveals that, in 90% of the distribution, the average spread is approximately 1%. At 95% it gets to 10.35%. Therefore, to reduce possible concerns that the results could be affected by these cases, I repeat the analysis by eliminating observations when the spread is 10% or above. In Table VIII, I report results that are essentially identical to those reported in Table II.

4.3. The Role of Political Corruption

Recent evidence suggests that political corruption could influence cash holdings at the firm level. For instance, on a sample of US companies, Jayakody et al. (2023) report that companies operating in environments with higher corruption levels find it beneficial to accumulate more cash. Accordingly, I collect the indicator of "Control of Corruption Estimate" from the World Bank.[8] This indicator "captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as capture of the state by elites and private interests." The index gives each country/year a score in units of a standard normal distribution, ranging from approximately -2.5 to 2.5 where higher values indicate lower levels of corruption The variable is available for the entire time series of my studies and for all countries. In broad strokes, this metric reflects the degree to which public authority is misused for personal benefit. According to Hamilton and Hammer (2018), this is one of the most reliable indicators of political corruption at the national level. I then include this variable as a control in the analysis. The results of this test are presented in Table IX. The variable for political corruption is never significant, but more importantly, my tests on the role of large shareholder diversification on cash holdings are essentially unaltered. For extra robustness, I also repeat the test by dropping countries with worse corruption levels, as indicated by the values of the index below zero. This excludes approximately 15% of the sample. The results in Table X largely mirror those of all the previous tests.

[8] https://databank.worldbank.org/source/worldwide-governance-indicators?l=en#

4.4. The Role of Disclosure Requirements

In this final series of robustness tests, I try to reduce concerns that heterogeneity in the disclosure requirements in my sample may introduce a significant amount of noise in my tests.

First, I exclude countries where the disclosure of financial statements is voluntary (Bosnia and Herzegovina, Romania, Russia, and Switzerland). The voluntary nature of the disclosure in these countries may raise self-selection concerns that may affect the main results, so this test shows that my results are not affected by this issue. Second, I further exclude countries where compliance with the disclosure requirements is either low (Portugal, Germany) or undefined (Malta and the Slovak Republic).

The results from these tests are reported in Tables XI and XII, respectively, and clearly indicate that the results are robust to heterogeneity in disclosures in different countries.

5. Conclusions

Several recent studies document that more diversified firms hold less cash (Duchin, 2010; Subramaniam et al., 2011; Tong, 2011; Bakke and Gu, 2017). A common thread of this literature is the focus on *segment* diversification at the *firm* level. In this study, I offer a different and original perspective by focusing on diversification at the *shareholder* level rather than at the *firm* level. I argue that large shareholders with less-diversified portfolios

are more concerned with firms' cash flow variability. Consequently, a higher level of cash holding at the firm level would reduce her exposure to this risk. Ceteris paribus, companies controlled by well-diversified shareholders are expected to accumulate less cash than those controlled by less-diversified shareholders.

I report a battery of tests, all of which corroborate my hypotheses. Shareholder-level portfolio diversification is not only statistically significant across all tests but also economically meaningful. In robust multivariate regression testing, measuring the economic impact in a conventional way, my tests indicate that one standard deviation increase in portfolio diversification (as measured by *Ln No. Firms*) results in an average 22% decrease in cash holdings relative to the mean. Results are robust when using the alternative proxies for portfolio diversification. Moreover, the results are robust to including a more conventional proxy for firm-level segment diversification.

Overall, my results highlight a robust and previously undocumented relation between the diversification of the largest shareholder and a firm's decision to accumulate cash. This negative relation does not seem to be driven by endogeneity or other potentially confounding effects such as the presence of dual-class shares, tunnelling, political corruption, or heterogeneity in disclosure requirements across courtiers. Economically, diversification of the largest firm owner is at least as important for determining the level of cash holdings as diversification of the firm's business across sectors is, and these effects complement each other.

References

Asker, J., Farre-Mensa, J., Ljungqvist, A., 2015. Corporate Investment and Stock Market Listing: A Puzzle? Rev. Financ. Stud. 28, 342–390. https://doi.org/10.1093/rfs/hhu077

Bakke, T.-E., Gu, T., 2017. Diversification and cash dynamics. Journal of Financial Economics 123, 580–601. https://doi.org/10.1016/j.jfineco.2016.12.008

Bates, T.W., Kahle, K.M., Stulz, R.M., 2009. Why Do U.S. Firms Hold So Much More Cash than They Used To? The Journal of Finance 64, 1985–2021. https://doi.org/10.1111/j.1540-6261.2009.01492.x

Bodnaruk, A., Kandel, E., Massa, M., Simonov, A., 2008. Shareholder Diversification and the Decision to Go Public. Rev. Financ. Stud. 21, 2779–2824. https://doi.org/10.1093/rfs/hhm036

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. Journal of the American Statistical Association 90, 443–450. https://doi.org/10.1080/01621459.1995.10476536

Campa, J.M., Kedia, S., 2002. Explaining the Diversification Discount. The Journal of Finance 57, 1731–1762.

Claessens, S., Djankov, S., Lang, L.H.P., 2000. The separation of ownership and control in East Asian Corporations. Journal of Financial Economics 58, 81–112. https://doi.org/10.1016/S0304-405X(00)00067-2

Duchin, R., 2010. Cash Holdings and Corporate Diversification. The Journal of Finance 65, 955–992. https://doi.org/10.1111/j.1540-6261.2010.01558.x

Faccio, M., Lang, L.H.P., 2002. The ultimate ownership of Western European corporations. Journal of Financial Economics 65, 365–395. https://doi.org/10.1016/S0304-405X(02)00146-0

Faccio, M., Marchica, M.-T., Mura, R., 2011. Large Shareholder Diversification and Corporate Risk-Taking. Rev. Financ. Stud. 24, 3601–3641. https://doi.org/10.1093/rfs/hhr065

Fernandes, N., Gonenc, H., 2016. Multinationals and cash holdings. Journal of Corporate Finance.

Gao, H., Harford, J., Li, K., 2013. Determinants of corporate cash policy: Insights from private firms. Journal of Financial Economics 109, 623–639. https://doi.org/10.1016/j.jfineco.2013.04.008

Gu, T., 2017. U.S. multinationals and cash holdings. Journal of Financial Economics 125, 344–368. https://doi.org/10.1016/j.jfineco.2017.05.007

Hall, T., Mateus, C., Mateus, I.B., 2014. What determines cash holdings at privately held and publicly traded firms? Evidence from 20 emerging markets. International Review of Financial Analysis 33, 104–116. https://doi.org/10.1016/j.irfa.2013.11.002

Hamilton, A., Hammer, C., 2018. Can We Measure the Power of the Grabbing Hand? A Comparative Analysis of Different Indicators of Corruption. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-8299

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47, 153–161. https://doi.org/10.2307/1912352

Jayakody, S., Morelli, D., Oberoi, J., 2023. Political uncertainty, corruption, and corporate cash holdings. Journal of Corporate Finance 82, 102447. https://doi.org/10.1016/j.jcorpfin.2023.102447

John, K., Litov, L., Yeung, B., 2008. Corporate Governance and Risk-Taking. The Journal of Finance 63, 1679–1728. https://doi.org/10.1111/j.1540-6261.2008.01372.x

Kai, L., Prabhala, N.R., 2007. Self-Selection Models in Corporate Finance**We thank N.K. Chidambaran, Craig Doidge, Espen Eckbo, Andrew Karolyi, Gordon Phillips, Vojislav Maksimovic, Jeffrey Smith, and Xinlei Zhao without implicating them for any errors or omissions, which remain ours. Li acknowledges the financial support from the Social Sciences and Humanities Research Council of Canada, and the W.M. Young Chair in Finance from the Sauder School of Business at UBC. Li also wishes to thank the support and hospitality of the MIT Sloan School of Management where she completed most of her work on this chapter., in: Handbook of Empirical Corporate Finance. Elsevier, pp. 37–86. https://doi.org/10.1016/B978-0-444-53265-7.50016-0

Laeven, L., Levine, R., 2009. Bank governance, regulation and risk taking. Journal of Financial Economics.

Laeven, L., Levine, R., 2007. Is there a diversification discount in financial conglomerates?$. Journal of Financial Economics.

Lauterbach, B., Pajuste, A., 2015. The long-term valuation effects of voluntary dual class share unifications. Journal of Corporate Finance 31, 171–185. https://doi.org/10.1016/j.jcorpfin.2015.02.004

Lyandres, E., Marchica, M.-T., Michaely, R., Mura, R., 2019. Owners' Portfolio Diversification and Firm Investment. The Review of Financial Studies 32, 4855–4904. https://doi.org/10.1093/rfs/hhz050

Maury, B., Pajuste, A., 2011. Private Benefits of Control and Dual-Class Share Unifications: PRIVATE BENEFITS OF CONTROL AND SHARE UNIFICATIONS. Manage. Decis. Econ. 32, 355–369. https://doi.org/10.1002/mde.1538

Mortal, S., Nanda, V., Reisel, N., 2020. Why do private firms hold less cash than public firms? International evidence on cash holdings and borrowing costs. Journal of Banking & Finance 113, 105722. https://doi.org/10.1016/j.jbankfin.2019.105722

Mortal, S., Reisel, N., 2013. Capital Allocation by Public and Private Firms. J. Financ. Quant. Anal. 48, 77–103. https://doi.org/10.1017/S0022109013000057

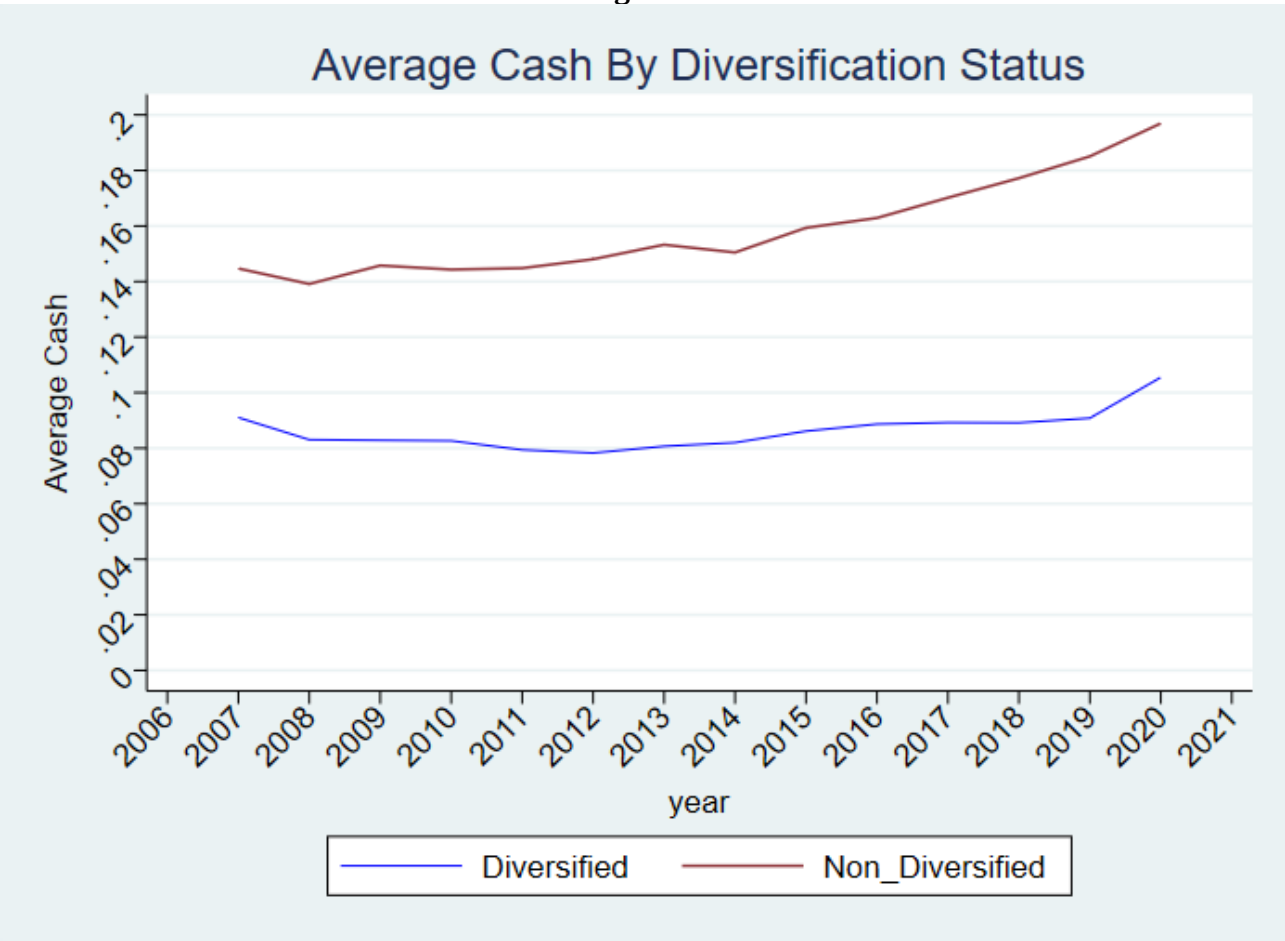
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55. https://doi.org/10.1093/biomet/70.1.41

Staiger, D., Stock, J.H., 1997. Instrumental Variables Regression with Weak Instruments. Econometrica 65, 557–586. https://doi.org/10.2307/2171753

Subramaniam, V., Tang, T.T., Yue, H., Zhou, X., 2011. Firm structure and corporate cash holdings. Journal of Corporate Finance 17, 759–773. https://doi.org/10.1016/j.jcorpfin.2010.06.002

Tong, Z., 2011. Firm diversification and the value of corporate cash holdings. Journal of Corporate Finance 17, 741–758. https://doi.org/10.1016/j.jcorpfin.2009.05.001

Figure 1



This figure plots the average cash holdings of firms whose large shareholders are either diversified or non-diversified during the period 2007-2020. Cash holding is the ratio of cash and cash equivalent to total assets. A shareholder is defined as non-diversified if she has only one equity position, while a diversified shareholder has two or more equity positions in her portfolio.

Table I Descriptive statistics

	Obs	Mean	Median	Min	Max	Std. Dev.
			Panel A: Investor Level			
Diversification	1,842,468	9.171	3	1.000	1267	42.346
Ln No.Firms	1,842,468	1.225	1.099	0.000	7.144	1.122
1-Herfindahl	1,821,158	0.530	0.667	0.000	0.998	0.348
Cash Flow Rights	1,842,468	60.032	51	0.000	100	32.246
Voting Rights	1,842,468	61.208	51	0.000	100	31.231
Spread	1,842,468	1.176	0.000	0.000	71.901	4.368
			Panel B: Firm Level			
Cash Holding	2,889,134	0.099	0.030	0.000	0.980	0.163
Private Company	2,889,134	0.989	1.000	0.000	1.000	0.104
Growth Opportunities	2,889,134	0.160	0.023	-18.152	7.106	0.813
Cash Flow	2,889,134	0.051	0.046	-4.111	1.500	0.224
Cash Flow Volatility	2,889,134	0.438	0.387	0.000	3.197	0.251
Age	2,889,134	23.009	19.000	1.000	98.000	18.098
Ln(1+Age)	2,889,134	2.812	2.944	0.000	4.585	0.881
Age	2,889,134	23.009	19.000	1.000	98.000	18.098
Ln(TA)	2,889,134	16.506	16.268	0.046	26.076	1.221
Leverage	2,889,134	0.647	0.663	0.000	2.200	0.349
Sector Diversification	2,889,134	1.768	1.000	1.000	35.000	1.097

Panel A of this table reports the descriptive statistics of the variables at the controlling-owner level. Diversification is the total number of firms in which a company's largest (ultimate) shareholder holds shares, directly or indirectly. Ln No. Firms is the natural log of the level variable described above. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. The transformation 1-Herfindal ensures that the interpretation is consistent with that of the Ln No. Firms. Cash Flow Rights measure the cash flow rights of the largest ultimate shareholder. Voting Rights measure the voting rights of the largest ultimate shareholder. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Panel B reports descriptive statistics at the firm level. Cash holding is the ratio of cash and cash equivalent to total assets. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level; Age is the number of years since incorporation; and Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Sector Diversification measures the number of business segments a firm reports. Ln(Sector Diversification) represents the natural log transformation of the same variable.

Table II OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.013***		-0.020***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.095***		-0.060***		-0.070***
		(0.000)		(0.000)		(0.000)
Private Company	-0.004***	-0.004***	-0.018***	-0.018***	-0.023***	-0.024***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.004***	0.004***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.012***	0.015***	0.010***	0.010***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.008***	0.006***	-0.001	-0.001	-0.002***	-0.002**
	(0.000)	(0.000)	(0.282)	(0.391)	(0.001)	(0.012)
Ln(1+Age)	0.001***	0.001***	0.004***	0.004***	-0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.199)	(0.284)
Ln(TA)	-0.013***	-0.014***	-0.013***	-0.014***	-0.021***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.051***	-0.051***	-0.019***	-0.019***	-0.024***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.642)	(0.000)
Ln (Sector Diversification)	0.002***	0.002***	-0.000	-0.000		
	(0.003)	(0.000)	(0.898)	(0.839)		
Constant	0.365***	0.424***	0.380***	0.376***	0.506***	0.530***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,889,134	2,858,737	2,889,134	2,858,737	2,889,134	2,858,737
R-squared	0.088	0.096	0.585	0.588	0.588	0.592
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-15.12		-22.67		-13.32	
EI (1-Herfindal)		-33.49		-21.07		-24.72
EI Ln (Sector Diversification)	0.830	0.982	-0.0431	-0.0683		

This table reports OLS regression results. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table III: IV Regression based on diversification

	(1)	(2) First	(3)	(4) First	(5)	(6) First	(7)	(8) First
Ln No.Firms	-0.053***				-0.013***			
	(0.000)				(0.000)			
IV (Ln No.Firms)		0.0196***				0.192***		
		(0.000)				(0.000)		
(1-Herfindahl)			-0.146***				-0.027***	
			(0.000)				(0.001)	
IV (1-Herfindahl)				0.0531***				0.162***
				(0.000)				(0.000)
Private Company	-0.018***	0.0136***	-0.019***	0.00417***	-0.024***	-0.0504	-0.024***	-0.0116***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.193)	(0.000)	(0.008)
Growth Options	0.002***	-0.00438***	0.002***	-0.00126***	0.002***	-0.00642***	0.002***	-0.00235***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.010***	0.000597	0.010***	0.00240***	0.007***	-0.0538***	0.007***	-0.00699***
	(0.000)	(0.690)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Vol.	-0.000	0.0146***	-0.000	0.00737***	-0.002***	0.0323***	-0.002***	0.0132***
	(0.582)	(0.000)	(0.843)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Ln(1+Age)	0.003***	-0.00834***	0.004***	-0.00170***	-0.001	0.0880***	-0.001**	0.0109***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.268)	(0.000)	(0.040)	(0.000)
Ln(TA)	-0.013***	0.0107***	-0.013***	0.00236***	-0.021***	0.0676***	-0.022***	0.0142***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.018***	0.00949***	-0.019***	-0.00478***	-0.024***	-0.0663***	-0.021***	-0.0234***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	0.000***	0.00657***	0.000***	0.00118***	0.000	0.0422***	-0.000***	0.00490***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.265)	(0.000)	(0.000)	(0.000)
Ln(Sector Diver.)	0.000	0.00503**	0.000	0.000217				
	(0.734)	(0.033)	(0.951)	(0.632)				
Observations	2,854,864	2,854,864	2,824,381	2,824,381	2,854,864	2,854,864	2,824,381	2,824,381
Partial R-sq		0.363		0.128		1.012		0.177
F-Test		1329		1866		6229		4289
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO	NO	NO
Investor FE	YES	YES	YES	YES	NO	NO	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES	YES	YES

This table reports the results of the Instrumental Variable regression. Odd-numbered models represent the second stage, while even-numbered models report the first stage. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. The instruments

for these variables are calculated as the average portfolio diversification (either Ln No.Firms or 1-Herfindahl) of large shareholders of all the other companies in the same country year and industry. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Sector Diversification measures the number of business segments a firm reports. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. For brevity, I report only the Investor Fixed Effects and Firm Fixed Effects models. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table IV: IV Regressions based on scrambled diversification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		First		First		First		First
Ln No.Firms	-0.011**				-0.008***			
	(0.036)				(0.000)			
IV (Ln No.Firms)		0.0198***				0.142***		
		(0.000)				(0.000)		
(1-Herfindahl)			-0.132***				-0.112***	
			(0.007)				(0.000)	
IV(1-Herfindahl)				0.0175***				0.0662***
				(0.000)				(0.000)
Private Company	-0.019***	0.0132***	-0.019***	0.00231***	-0.023***	-0.134***	-0.024***	-0.0133***
	(0.000)	(0.003)	(0.000)	(0.007)	(0.000)	(0.000)	(0.000)	(0.004)
Growth Options	0.003***	-0.00486***	0.002***	-0.00158***	0.003***	-0.00997***	0.003***	-0.00313***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.010***	-4.53e-05	0.011***	0.00230***	0.007***	-0.0571***	0.007***	-0.00722***
	(0.000)	(0.980)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Vol.	-0.000	0.0229***	0.001	0.00907***	-0.002***	0.0524***	-0.000	0.0189***
	(0.556)	(0.000)	(0.493)	(0.000)	(0.005)	(0.000)	(0.631)	(0.000)
Ln(1+Age)	0.003***	-0.00119*	0.003***	-0.000913***	-0.000	0.0967***	0.001*	0.0182***
	(0.000)	(0.053)	(0.000)	(0.000)	(0.431)	(0.000)	(0.064)	(0.000)
Ln(TA)	-0.014***	0.00902***	-0.014***	0.00172***	-0.022***	0.0637***	-0.022***	0.0127***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.022***	0.0152***	-0.022***	-0.00312***	-0.028***	-0.0449***	-0.028***	-0.0167***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	0.000***	0.00588***	0.000***	0.000892***	-0.000**	0.0358***	0.000	0.00391***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.014)	(0.000)	(0.244)	(0.000)
Ln(Sector Diver.)	-0.001	0.00787***	-0.001	0.000770				
	(0.210)	(0.003)	(0.253)	(0.131)				
Observations	2,060,990	2,060,990	2,039,570	2,039,570	2,060,990	2,060,990	2,039,570	2,039,570
Partial R-sq		0.356		0.106		0.983		0.154
F-Test		3059		376.8		18237		2666
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO	NO	NO
Investor FE	YES	YES	YES	YES	NO	NO	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES	YES	YES

This table reports the results of the Instrumental Variable regression. Odd-numbered models represent the second stage, while even-numbered models report the first stage. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or

indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. The instruments for these variables are calculated as the average portfolio diversification of large shareholders of all other companies in the same country year and industry, calculated for a matching firm in a neighboring country. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Sector Diversification measures the number of business segments a firm reports. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. The economic significance of the portfolio diversification variables is reported below the p-values. For brevity, I report only the Investor Fixed Effects and Firm Fixed Effects models. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table V: Heckman Correction Models

	(1) No Excl. Restriction	(2) No Excl. Restriction	(3) Geo Distance	(4) Geo Distance	(5) No Excl. Restriction	(6) No Excl. Restriction	(7) Geo Distance	(8) Geo Distance
Ln No.Firms	-0.015***		-0.014***		-0.010***		-0.010***	
	(0.000)		(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.069***		-0.069***		-0.064***		-0.064***
		(0.000)		(0.000)		(0.000)		(0.000)
Lambda	-0.016***	0.002	-0.016***	0.002	-0.022***	-0.004***	-0.022***	-0.004***
	(0.000)	(0.105)	(0.000)	(0.125)	(0.000)	(0.004)	(0.000)	(0.002)
Private Company	-0.015***	-0.016***	-0.015***	-0.016***	-0.018***	-0.019***	-0.018***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.002)	(0.004)	(0.002)
Growth Opportunities	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.007***	0.007***	0.007***	0.007***	0.005***	0.005***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.001)	(0.003)
Cash Flow Volatility	0.001	0.002	0.002	0.002	-0.001	0.000	-0.000	0.000
	(0.273)	(0.104)	(0.260)	(0.105)	(0.725)	(0.794)	(0.752)	(0.792)
Ln(1+Age)	0.003***	0.003***	0.003***	0.003***	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.270)	(0.328)	(0.269)	(0.329)
Ln(TA)	-0.017***	-0.016***	-0.017***	-0.016***	-0.025***	-0.025***	-0.025***	-0.025***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.023***	-0.023***	-0.023***	-0.023***	-0.029***	-0.027***	-0.029***	-0.027***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	0.000	0.000***	-0.000	0.000***	-0.000**	-0.000***	-0.000**	-0.000***
	(0.988)	(0.002)	(0.984)	(0.002)	(0.015)	(0.000)	(0.014)	(0.000)
Ln(Sector Diversification)	-0.002	-0.002	-0.002	-0.002				
	(0.150)	(0.175)	(0.148)	(0.176)				
Constant	0.435***	0.440***	0.435***	0.440***	0.576***	0.595***	0.575***	0.595***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	583,801	578,192	583,801	578,192	583,801	578,192	583,801	578,192
R-squared	0.586	0.588	0.586	0.588	0.684	0.686	0.684	0.686
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO	NO	NO
Investor FE	YES	YES	YES	YES	NO	NO	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES	YES	YES

This table reports the results obtained using the Heckman treatment effects model. For brevity, I report only the Investor Fixed Effects and Firm Fixed Effects models. Models (1), (2), (5), and (6) are produced using a first-stage probit, where no exclusion restriction is included. Models (3), (4), (7), and (8) are produced via a first-stage probit, where the average number of companies located within a 50 miles radius of each large shareholder in each country in each year is used as an exclusion restriction in the first stage. Lambda is calculated from the predicted values of the first-stage probit regressions. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level; Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Sector Diversification measures the number of business segments a firm reports. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table VI: Propensity Score Matching

	Observations	Diversified	Undiversified	p-value of diff.
Cash Holding	75,904	0.074	0.174	0.000
P-Score	75,904	0.701	0.701	0.897
Private Company	75,904	0.999	0.999	1.000
Growth Opportunities	75,904	0.349	0.349	0.984
Cash Flow	75,904	0.037	0.037	0.945
Cash Flow Volatility	75,904	0.453	0.453	0.850
Age	75,904	15.839	15.839	1.000
Ln(TA)	75,904	16.130	16.129	0.933
Leverage	75,904	0.742	0.744	0.580
Spread	75,904	0.001	0.001	0.802
Sector Diversification	75,904	1.546	1.546	1.000

This table reports the results obtained using a propensity score matching procedure. In the first step, I run a logit model on the probability of a firm having a dominant shareholder that is either undiversified (only one equity position) or diversified (more than one equity position in the portfolio). Matching is performed within each country, industry, year, public/private status, age, and segment diversification. In the logit model, I then add all the non-discrete controls. To ensure the quality of the matching, I impose a maximum difference in the propensity score (caliper) between the treated and control firms to not exceed 0.001 in absolute value. Cash holding is the ratio of cash and cash equivalent to total assets. The P-score is the propensity score, which represents the estimated probability of treatment assignment. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Age is defined as the number of years since incorporation. Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Sector Diversification measures the number of business segments a firm reports. Discrete variables are reported for completeness although they are matched within each discrete value. P-values of tests of differences in means between the treated and control groups are reported in the last column.

Table VII: Excluding Counties with a Higher Fraction of Dual-Share Classes

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.013***		-0.021***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.100***		-0.064***		-0.075***
		(0.000)		(0.000)		(0.000)
Private Company	-0.004***	-0.006***	-0.015***	-0.015***	-0.016***	-0.018***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.003***	0.003***	0.002***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.008***	0.009***	0.008***	0.008***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.002***	0.001**	-0.001*	-0.001*	-0.001	-0.000
	(0.002)	(0.046)	(0.056)	(0.094)	(0.258)	(0.942)
Ln(1+Age)	0.000**	0.001***	0.005***	0.005***	0.003***	0.003***
	(0.017)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(TA)	-0.013***	-0.015***	-0.014***	-0.014***	-0.021***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.044***	-0.043***	-0.017***	-0.016***	-0.023***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.635)	(0.000)
Ln (Sector Diversification)	0.001	0.001**	0.001	0.001		
	(0.138)	(0.039)	(0.211)	(0.207)		
Constant	0.384***	0.440***	0.387***	0.380***	0.500***	0.523***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,915,897	1,893,664	1,915,897	1,893,664	1,915,897	1,893,664
R-squared	0.089	0.098	0.592	0.596	0.591	0.595
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-15.94		-25.45		-14.50	
EI (1-Herfindal)		-35.42		-22.55		-26.45
EI Ln (Sector Diversification)	0.470	0.657	0.478	0.483		

This table reports OLS regression results, where I exclude the seven European countries with more prevalent use of dual-class shares: Denmark, Finland, Germany, Italy, Norway, Sweden, and Switzerland. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table VIII: Excluding Firms with A Higher Risk of Tunnelling

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.014***		-0.021***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.095***		-0.060***		-0.070***
		(0.000)		(0.000)		(0.000)
Private Company	-0.004***	-0.003***	-0.019***	-0.019***	-0.024***	-0.025***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.004***	0.004***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.011***	0.014***	0.009***	0.010***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.007***	0.006***	-0.001	-0.001	-0.002***	-0.002**
	(0.000)	(0.000)	(0.272)	(0.415)	(0.002)	(0.032)
Ln(1+Age)	0.000*	0.001***	0.004***	0.004***	-0.000	-0.000
	(0.091)	(0.000)	(0.000)	(0.000)	(0.414)	(0.846)
Ln(TA)	-0.013***	-0.014***	-0.013***	-0.014***	-0.021***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.053***	-0.052***	-0.019***	-0.018***	-0.024***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	0.001***	-0.001***	0.001***	0.000***	0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)
Ln (Sector Diversification)	0.002***	0.002***	0.000	0.000		
	(0.003)	(0.001)	(0.729)	(0.782)		
Constant	0.368***	0.424***	0.381***	0.377***	0.502***	0.526***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,623,589	2,594,280	2,623,589	2,594,280	2,623,589	2,594,280
R-squared	0.087	0.095	0.596	0.599	0.592	0.597
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-15.34		-23.18		-13.44	
EI (1-Herfindal)		-32.96		-20.70		-24.22
EI Ln (Sector Diversification)	0.862	1.006	0.124	0.0989		

This table reports the OLS regression results, where I exclude cases in which voting rights exceed cash flow rights by 10% or more. This corresponds to about the top 5% of the total sample. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table IX: Controlling for the Levels of Political Corruption

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.013***		-0.020***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.095***		-0.060***		-0.070***
		(0.000)		(0.000)		(0.000)
Control Corruption Estimate	0.004	0.005	0.028	0.027	-0.005	-0.010
	(0.902)	(0.886)	(0.198)	(0.222)	(0.803)	(0.646)
Private Company	-0.004***	-0.004***	-0.018***	-0.018***	-0.023***	-0.024***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.004***	0.004***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.012***	0.015***	0.010***	0.010***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.008***	0.006***	-0.001	-0.001	-0.002***	-0.002**
	(0.000)	(0.000)	(0.281)	(0.390)	(0.001)	(0.012)
Ln(1+Age)	0.001***	0.001***	0.004***	0.004***	-0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.199)	(0.284)
Ln(TA)	-0.013***	-0.014***	-0.013***	-0.014***	-0.021***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.051***	-0.051***	-0.019***	-0.019***	-0.024***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.642)	(0.000)
Ln (Sector Diversification)	0.002***	0.002***	-0.000	-0.000		
	(0.003)	(0.000)	(0.899)	(0.840)		
Constant	0.362***	0.420***	0.355***	0.353***	0.510***	0.539***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,889,134	2,858,737	2,889,134	2,858,737	2,889,134	2,858,737
R-squared	0.088	0.096	0.585	0.588	0.588	0.592
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-15.12		-22.67		-13.32	
EI (1-Herfindal)		-33.49		-21.07		-24.72
EI Ln (Sector Diversification)	0.830	0.982	-0.0425	-0.0677		

This table reports the OLS regression results, where I include a proxy to control for political-corruption. Control Corruption Estimate captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5 where higher values indicate lower levels of corruption (https://databank.worldbank.org/source/worldwide-governance-indicators?l=en#). The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table X: Excluding Countries with High Levels of Political Corruption Levels

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.013***		-0.018***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.087***		-0.052***		-0.062***
		(0.000)		(0.000)		(0.000)
Private Company	-0.007***	-0.006***	-0.027***	-0.027***	-0.033***	-0.032***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.004***	0.004***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.016***	0.019***	0.011***	0.012***	0.008***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.011***	0.009***	-0.000	-0.000	-0.001*	-0.001*
	(0.000)	(0.000)	(0.878)	(0.879)	(0.059)	(0.088)
Ln(1+Age)	0.002***	0.003***	0.004***	0.004***	-0.004***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(TA)	-0.013***	-0.016***	-0.014***	-0.014***	-0.021***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.058***	-0.059***	-0.019***	-0.019***	-0.026***	-0.026***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.122)	(0.000)
Ln (Sector Diversification)	-0.002***	-0.002***	-0.001	-0.001		
	(0.001)	(0.001)	(0.226)	(0.183)		
Constant	0.380***	0.447***	0.390***	0.385***	0.534***	0.556***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,465,856	2,445,450	2,465,856	2,445,450	2,465,856	2,445,450
R-squared	0.100	0.104	0.579	0.581	0.601	0.603
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-14.77		-20.15		-13.11	
EI (1-Herfindal)		-29.69		-17.66		-21.10
EI Ln (Sector Diversification)	-1.015	-1.013	-0.437	-0.481		

This table reports the OLS regression results, where countries are dropped when the "Control Corruption Estimate" index is negative. As discussed above the index ranges from approximately -2.5 to 2.5 where higher values indicate lower levels of corruption (https://databank.worldbank.org/source/worldwide-governance-indicators?l=en#). This corresponds to a decrease of approximately 15% in the total sample. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table XI: Excluding Countries with Voluntarily Disclosure

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.013***		-0.018***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.077***		-0.047***		-0.057***
		(0.000)		(0.000)		(0.000)
Private Company	-0.006***	-0.005***	-0.023***	-0.023***	-0.025***	-0.025***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.004***	0.004***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.015***	0.018***	0.011***	0.012***	0.009***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.010***	0.009***	-0.000	-0.000	-0.002***	-0.001**
	(0.000)	(0.000)	(0.579)	(0.773)	(0.008)	(0.044)
Ln(1+Age)	0.002***	0.002***	0.004***	0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(TA)	-0.013***	-0.015***	-0.014***	-0.014***	-0.021***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.054***	-0.054***	-0.018***	-0.018***	-0.024***	-0.023***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.774)	(0.000)
Ln (Sector Diversification)	-0.001**	-0.001*	-0.000	-0.000		
	(0.042)	(0.052)	(0.586)	(0.522)		
Constant	0.374***	0.425***	0.386***	0.376***	0.524***	0.540***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,621,330	2,603,918	2,621,330	2,603,918	2,621,330	2,603,918
R-squared	0.092	0.094	0.575	0.577	0.590	0.592
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-14.68		-20.32		-13.06	
EI (1-Herfindal)		-27.18		-16.77		-20.18
EI Ln (Sector Diversification)	-0.579	-0.556	-0.187	-0.221		

This table reports the OLS regression results, where I exclude countries where the disclosure of financial statements is voluntary. These are Bosnia and Herzegovina, Romania, Russia, and Switzerland. The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table XII: Excluding Countries with Low Disclosure Compliance

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.013***		-0.019***		-0.012***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.078***		-0.049***		-0.059***
		(0.000)		(0.000)		(0.000)
Private Company	-0.007***	-0.007***	-0.022***	-0.023***	-0.024***	-0.024***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.004***	0.004***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.014***	0.017***	0.011***	0.011***	0.008***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.009***	0.007***	-0.001**	-0.001*	-0.003***	-0.002***
	(0.000)	(0.000)	(0.038)	(0.065)	(0.000)	(0.001)
Ln(1+Age)	0.002***	0.001***	0.004***	0.004***	-0.003***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(TA)	-0.013***	-0.015***	-0.014***	-0.014***	-0.021***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.050***	-0.050***	-0.017***	-0.016***	-0.022***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.727)	(0.000)
Ln (Sector Diversification)	-0.002***	-0.002***	-0.000	-0.000		
	(0.000)	(0.000)	(0.634)	(0.572)		
Constant	0.379***	0.430***	0.387***	0.377***	0.525***	0.542***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,311,163	2,295,489	2,311,163	2,295,489	2,311,163	2,295,489
R-squared	0.089	0.092	0.567	0.569	0.580	0.582
Year FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	NO	NO
SIC4 FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-15.12		-21.44		-13.82	
EI (1-Herfindal)		-28.37		-17.71		-21.42
EI Ln (Sector Diversification)	-1.259	-1.179	-0.182	-0.216		

This table reports OLS regression results where I exclude countries with voluntary disclosure of financial statements as above and exclude countries with low compliance with the disclosure requirements (Portugal, Germany) or those with undefined disclosure requirements (Malta, Monaco, and Slovak Republic). The dependent variable is corporate cash holdings, defined as the ratio of cash and cash equivalent to total assets. Ln No. Firms is the natural log of the total number of firms in which a company's largest ultimate shareholder (e.g., the ultimate shareholder controlling the largest fraction of voting rights in the firm) holds shares directly or indirectly in a given year. The Herfindahl Index is the sum of the squared values of the weight of each investment in the largest shareholder's portfolio. Private Company is a dummy variable that takes the value of one when the firm is privately held and zero when the firm is listed on a stock exchange. Growth Opportunities is the annual growth rate of total assets. Cash Flow is the ratio of income plus depreciation to total assets. Cash Flow Volatility is the standard deviation of cash flows at the country-year-industry level. Ln(1+Age) is the natural log of (1 + number of years since incorporation). Ln(TA) is the natural log of total assets expressed in 1999 prices. Leverage is the ratio of total debt to total assets. Spread measures the difference between the voting rights and cash flow rights of the ultimate owner. Ln(Sector Diversification) measures the natural log of the number of business segments reported by a firm. EI stands for Economic Impact, which is calculated by multiplying the estimated coefficient of the variable by one standard deviation of the same variable, and the product is then divided by the median of the dependent variable. All regressions include year fixed effects. P-values adjusted for heteroscedasticity are reported in brackets below the coefficients. Robust pval in parentheses *** p<0.01, ** p<0.05, * p<0.1

LARGE SHAREHOLDER DIVERSIFICATION AND CORPORATE CASH HOLDING

Internet Appendix

In this section, I report on a series of additional tables and further robustness tests performed on the data.

Table A1 reports the breakdown of sample coverage by country.

In Table A2, I replicate my main tests but use growth in sales as a proxy for growth opportunities instead of growth in total assets.

In Tables A3, A4, and A5, three variations of the Heckman correction model are reported. More specifically, Table A3 reports the results where no exclusion restriction is imposed in the first stage, and the sample is not restricted to observations with data for geographical distance being available (as in Table V). In Tables A4 and A5, I base the measure of investable firms on measures of geographical distance of five miles (A4) and 25 miles (A5). As reported in footnote 7 above, the impact of the exclusion restriction is limited, and reporting a maximum of three decimals results in seemingly identical tables. For this reason, here I report six decimals.

Table A1

	Observations	Percentage of sample	Cumulative percentage of sample
ALBANIA	738	0.03	0.03
AUSTRIA	36,302	1.26	1.28
BELGIUM	82,574	2.86	4.14
BOSNIA HERZEGOVINA	6,983	0.24	4.38
BULGARIA	24,461	0.85	5.23
CROATIA	15,148	0.52	5.75
CYPRUS	1,271	0.04	5.80
CZECH REPUBLIC	34,912	1.21	7.01
DENMARK	55,804	1.93	8.94
ESTONIA	10,541	0.36	9.30
FINLAND	26,409	0.91	10.22
FRANCE	283,432	9.81	20.03
GERMANY	222,464	7.70	27.73
GREECE	34,849	1.21	28.93
HUNGARY	10,635	0.37	29.30
ICELAND	3,894	0.13	29.44
IRELAND	27,914	0.97	30.40
ITALY	613,440	21.23	51.63
KOSOVO	61	0.00	51.64
LATVIA	6,847	0.24	51.87
LITHUANIA	8,042	0.28	52.15
LUXEMBOURG	6,995	0.24	52.39
MALTA	3,567	0.12	52.52
MONTENEGRO	1,889	0.07	52.58
NETHERLANDS	37,322	1.29	53.87
MACEDONIA	2,939	0.10	53.98
NORWAY	84,240	2.92	56.89
POLAND	88,366	3.06	59.95
PORTUGAL	68,669	2.38	62.33
REPUBLIC OF MOLDOVA	234	0.01	62.34
ROMANIA	46,736	1.62	63.95
RUSSIAN FEDERATION	213,579	7.39	71.35
SERBIA	19,384	0.67	72.02
SLOVAKIA	15,467	0.54	72.55
SLOVENIA	12,566	0.43	72.99
SPAIN	320,771	11.10	84.09
SWEDEN	54,614	1.89	85.98
SWITZERLAND	506	0.02	86.00
TURKEY	41,247	1.43	87.42
UKRAINE	30,178	1.04	88.47
UNITEDKINGDOM	333,144	11.53	100.00
Total	2,889,134	100.00	

This table reports the distribution of observations in the sample, divided by country.

Table A2

	(1)	(2)	(3)	(4)	(5)	(6)
Ln No.Firms	-0.011***		-0.015***		-0.009***	
	(0.000)		(0.000)		(0.000)	
(1-Herfindahl)		-0.073***		-0.044***		-0.048***
		(0.000)		(0.000)		(0.000)
Private Company	0.002**	0.002*	-0.015***	-0.016***	-0.022***	-0.023***
	(0.035)	(0.055)	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.010)	(0.004)
Cash Flow	0.028***	0.032***	0.018***	0.020***	0.013***	0.014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	0.013***	0.011***	0.002**	0.002***	-0.001	-0.001
	(0.000)	(0.000)	(0.013)	(0.008)	(0.188)	(0.274)
Ln(1+Age)	0.000***	0.001***	0.002***	0.002***	-0.004***	-0.004***
	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(TA)	-0.011***	-0.013***	-0.011***	-0.011***	-0.015***	-0.015***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.069***	-0.072***	-0.026***	-0.027***	-0.027***	-0.028***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000***	-0.000***	0.000***	0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.341)	(0.000)
Ln (Sector Diversification)	-0.001	-0.000	-0.000	-0.000		
	(0.435)	(0.517)	(0.979)	(0.942)		
Constant	0.333***	0.383***	0.328***	0.324***	0.408***	0.422***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,911,115	1,894,837	1,911,115	1,894,837	1,911,115	1,894,837
R-squared	0.096	0.101	0.575	0.577	0.596	0.598
Year FE	YES	YES	YES	YES	YES	YES
SIC4 FE	YES	YES	YES	YES	NO	NO
Country FE	YES	YES	YES	YES	NO	NO
Investor FE	NO	NO	YES	YES	NO	NO
Firm FE	NO	NO	NO	NO	YES	YES
EI Ln No.Firms	-12.57		-17.33		-10.13	
EI (1-Herfindal)		-25.67		-15.29		-16.98
EI Ln (Sector Diversification)	-0.264	-0.220	-0.0108	-0.0297		

This table reports the OLS regression results, where the growth rate in total sales is used as a proxy for growth options in place of growth in total assets. As discussed in the text, variable sales contain many more missing data points, which leads to a loss of observation.

Table A3

VARIABLES	(1) No Excl. Restriction	(2) No Excl. Restriction	(3) No Excl. Restriction	(4) No Excl. Restriction
Ln No.Firms	-0.014***		-0.009***	
	(0.000)		(0.000)	
(1-Herfindahl)		-0.056***		-0.066***
		(0.000)		(0.000)
lambda	-0.016***	-0.002***	-0.022***	-0.002***
	(0.000)	(0.002)	(0.000)	(0.000)
Private Company	-0.017***	-0.018***	-0.020***	-0.024***
	(0.000)	(0.000)	(0.000)	(0.000)
Growth Opportunities	0.002***	0.002***	0.003***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.010***	0.010***	0.008***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow Volatility	-0.001**	-0.001	-0.003***	-0.002***
	(0.047)	(0.327)	(0.000)	(0.007)
Ln(1+Age)	0.004***	0.004***	-0.000	-0.000
	(0.000)	(0.000)	(0.290)	(0.271)
Ln(TA)	-0.014***	-0.014***	-0.023***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.019***	-0.019***	-0.025***	-0.022***
	(0.000)	(0.000)	(0.000)	(0.000)
Spread	-0.000	0.000***	-0.000***	-0.000***
	(0.149)	(0.000)	(0.000)	(0.000)
Ln (Sector Diversification)	-0.000	-0.000		
	(0.827)	(0.835)		
Constant	0.382***	0.376***	0.529***	0.531***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,889,122	2,858,725	2,889,122	2,858,725
R-squared	0.585	0.588	0.590	0.592
Year FE	YES	YES	YES	YES
Country FE	YES	YES	NO	NO
SIC4 FE	YES	YES	NO	NO
Investor FE	YES	YES	NO	NO
Firm FE	NO	NO	YES	YES

This table reports the Heckman treatment effects model, where no exclusion restriction is imposed in the first stage, and the sample is not restricted to observations with data for geographical distance being available.

Table A4

VARIABLES	(3) Geo DIstance	(4) Geo DIstance	(7) Geo DIstance	(8) Geo DIstance
Ln No.Firms	-0.014504***		-0.010246***	
	(0.000)		(0.000)	
(1-Herfindahl)		-0.069073***		-0.064187***
		(0.000)		(0.000)
Lambda	-0.016283***	0.002462	-0.021866***	-0.003702***
	(0.000)	(0.105)	(0.000)	(0.004)
Private Company	-0.014809***	-0.015886***	-0.017656***	-0.019348***
	(0.000)	(0.000)	(0.003)	(0.002)
Growth Opportunities	0.002532***	0.002538***	0.003286***	0.003337***
	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.007374***	0.007216***	0.005336***	0.005059***
	(0.000)	(0.000)	(0.001)	(0.003)
Cash Flow Volatility	0.001472	0.002196	-0.000500	0.000375
	(0.273)	(0.104)	(0.725)	(0.794)
Ln(1+Age)	0.003252***	0.003407***	0.001233	0.001103
	(0.000)	(0.000)	(0.270)	(0.328)
Ln(TA)	-0.017026***	-0.016354***	-0.025383***	-0.025122***
	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.023228***	-0.023176***	-0.028868***	-0.027467***
	(0.000)	(0.000)	(0.000)	(0.000)
Spread	0.000001	0.000152***	-0.000115**	-0.000273***
	(0.988)	(0.002)	(0.015)	(0.000)
Ln (Sector Diversification)	-0.001852	-0.001746		
	(0.150)	(0.175)		
Constant	0.435087***	0.439903***	0.575618***	0.594613***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	583,801	578,192	583,801	578,192
R-squared	0.586312	0.588129	0.683718	0.685678
Year FE	YES	YES	YES	YES
Country FE	YES	YES	NO	NO
SIC4 FE	YES	YES	NO	NO
Investor FE	YES	YES	NO	NO
Firm FE	NO	NO	YES	YES

This table reports the results obtained using the Heckman treatment effects model. For brevity, I report only the Investor Fixed Effects and Firm Fixed Effects models. Models (1), (2), (5), and (6) are omitted because they are produced via a first-stage probit where no exclusion restriction is included; thus, they are identical to those reported in Table V. Models (3), (4), (7), and (8) are produced via a first-stage probit, where the average number of companies located within a 5 miles radius of each large shareholder in each country in each year is used as an exclusion restriction in the first stage.

Table A5

VARIABLES	(3) Geo DIstance	(4) Geo DIstance	(7) Geo DIstance	(8) Geo DIstance
Ln No.Firms	-0.014501***		-0.010243***	
	(0.000)		(0.000)	
(1-Herfindahl)		-0.068986***		-0.063975***
		(0.000)		(0.000)
Lambda	-0.016308***	0.002421	-0.021894***	-0.003811***
	(0.000)	(0.111)	(0.000)	(0.003)
Private Company	-0.014785***	-0.015888***	-0.017620***	-0.019334***
	(0.000)	(0.000)	(0.004)	(0.002)
Growth Opportunities	0.002532***	0.002538***	0.003287***	0.003338***
	(0.000)	(0.000)	(0.000)	(0.000)
Cash Flow	0.007379***	0.007217***	0.005342***	0.005065***
	(0.000)	(0.000)	(0.001)	(0.003)
Cash Flow Volatility	0.001484	0.002193	-0.000487	0.000374
	(0.269)	(0.105)	(0.732)	(0.795)
Ln(1+Age)	0.003254***	0.003406***	0.001237	0.001102
	(0.000)	(0.000)	(0.269)	(0.329)
Ln(TA)	-0.017023***	-0.016356***	-0.025379***	-0.025129***
	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.023234***	-0.023175***	-0.028878***	-0.027467***
	(0.000)	(0.000)	(0.000)	(0.000)
Spread	0.000000	0.000151***	-0.000115**	-0.000275***
	(0.992)	(0.002)	(0.015)	(0.000)
Ln (Sector Diversification)	-0.001856	-0.001746		
	(0.150)	(0.176)		
Constant	0.435019***	0.439887***	0.575515***	0.594586***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	583,801	578,192	583,801	578,192
R-squared	0.586313	0.588128	0.683721	0.685679
Year FE	YES	YES	YES	YES
Country FE	YES	YES	NO	NO
SIC4 FE	YES	YES	NO	NO
Investor FE	YES	YES	NO	NO
Firm FE	NO	NO	YES	YES

This table reports the results obtained using the Heckman treatment effects model. For brevity, I report only the Investor Fixed Effects and Firm Fixed Effects models. Models (1), (2), (5), and (6) are omitted because they are produced via a first-stage probit where no exclusion restriction is included; thus, they are identical to those reported in Table V. Models (3), (4), (7), and (8) are produced via a first-stage probit, where the average number of companies located within a 25 miles radius of each large shareholder in each country in each year is used as an exclusion restriction in the first stage.