Journal of Financial Economics 90 (2008) 20–37

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

The only game in town: Stock-price consequences of local bias ☆

Harrison Hong [a,*], Jeffrey D. Kubik [b], Jeremy C. Stein [c]

[a] *Department of Economics, Princeton University, 26 Prospect Avenue, Princeton, NJ 08540-5296, USA*
[b] *Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020, USA*
[c] *Department of Economics, Harvard University, Littauer Room 209, Cambridge, MA 02138, USA*

ARTICLE INFO

Article history:
Received 24 April 2006
Received in revised form
19 November 2007
Accepted 29 November 2007
Available online 12 August 2008

JEL classification:
G11
G12

Keywords:
Local bias
Segmented markets

ABSTRACT

Theory suggests that, in the presence of local bias, the price of a stock should be decreasing in the ratio of the aggregate book value of firms in its region to the aggregate risk tolerance of investors in its region. Using data on U.S. states and Census regions, we find clear-cut support for this proposition. Most of the variation in the ratio of interest comes from differences across regions in aggregate book value per capita. Regions with low population density—e.g., the Deep South—are home to relatively few firms per capita, which leads to higher stock prices via an "only-game-in-town" effect.

1. Introduction

A number of recent papers show that investors have a tendency to be strongly locally biased in their portfolio choices. This bias shows up not only as a preference for domestic as opposed to foreign stocks (French and Poterba, 1991; Cooper and Kaplanis, 1994), but perhaps more strikingly, as a preference for those domestic stocks that are headquartered close by (Coval and Moskowitz, 1999; Grinblatt and Keloharju, 2000; Huberman, 2001). Both professional money managers and individual investors exhibit some degree of local bias, though it is substantially stronger among individuals (Zhu, 2003).

While the existence of within-country local bias now seems to be incontrovertible, there is little evidence regarding its equilibrium asset-pricing implications. In particular, we know of no work that attempts to relate the *level* of a firm's stock price to market conditions in its home locale. Yet basic theoretical considerations suggest that such a link should exist. The logic is most easily seen by considering an extreme case of local preference in which investors only purchase the stocks of companies headquartered in, say, their home state. In this case, each state is its own autarkic capital market, with a risk premium that is determined by the ratio of the total supply of shares in the state to the total risk tolerance of investors living in the state.[1]

In what follows, we investigate this hypothesis. We begin by constructing a variable we call RATIO, which for any given region at any point in time, is equal to the

☆ This research is supported by the National Science Foundation. We are grateful for the comments of Gur Huberman, Alok Kumar, Jake Thomas, Scott Weisbenner, the referee, and seminar participants at USC, Harvard Business School, the Yale School of Management, Princeton, the University of Iowa Behavioral Finance conference, and the Western Finance Association meetings.
* Corresponding author.
E-mail address: hhong@princeton.edu (H. Hong).

[1] This prediction is independent of the root cause of local bias—it could reflect feelings of loyalty among local investors, a lack of familiarity or awareness with respect to non-local stocks, endogenous information asymmetries (Van Nieuwerburgh and Veldkamp, 2007), or various transactional impediments.

0304-405X/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2007.11.006

aggregate book value of all firms headquartered in the region, divided by the aggregate income of all households living in the region. We do this both for individual states, as well as for the nine U.S. Census regions, though, for reasons we develop below, our primary focus is on the latter, coarser geographic level. Next, we run cross-sectional regressions of the log of a firm's market-to-book on RATIO, as well as on several controls (including return on equity, the ratio of R&D to sales, and industry and exchange dummies). Using a sample that runs from 1970 to 2005, we find that the RATIO variable has a negative impact on stock prices. In particular, if one goes from the Census region with the highest value of RATIO (the Middle Atlantic), to the region with the lowest value (the Deep South), holding all else equal, the implied increase in the stock price is about 7.9%.

Digging deeper, we find that our results are intimately connected to regional variation in population density. That is, regions with low population density—of which the Deep South is an example—tend to have low values of RATIO, which are associated with higher stock prices. This is because most of the variation in RATIO across regions is driven by the book value component, which is very sensitive to population density. Specifically, if one rewrites RATIO as total book value per capita divided by income per capita, it turns out that both per-capita variables are positively related to population density, but that book value per capita is much more responsive to density than is income per capita. In other words, in spite of low per-capita income, the Deep South is associated with higher stock prices because of an "only-game-in-town" effect: any one company headquartered there faces relatively little competition for local investors' dollars, because so few *other* companies are headquartered there.

Of course, the close correlation between population density and RATIO begs the question of whether the former is a variable that can legitimately be excluded from the right-hand side of the regression. In other words, are there reasons to think that population density might proxy for some other economic factor that exerts an independent influence on stock prices? One possibility is that regions with low population density have greater future growth prospects, and that it is these superior growth prospects that drive higher stock prices. We attempt to control for this possibility by adding to our baseline regressions a series of future-growth variables, such as future income growth at the regional level, as well as future sales growth and profitability at the firm level. None of these controls materially alter our basic results.

A related concern is that there may be compositional differences in the sorts of firms that are located in different parts of the country, and that are not fully captured by Standard Industrial Classification (SIC) code based industry dummies. For example, it might be that even within an SIC-defined industry, firms located in the Deep South use a different production technology—e.g., one that is more or less human-capital-intensive—than firms located in the Middle Atlantic, and that these technology differences are what lie behind the observed patterns in stock prices.

In an effort to come to grips with this issue, we re-run our baseline analysis on a subsample of electric utility firms. The premise here is that electric utilities are a relatively homogeneous set of firms with a common technology. Moreover, due to high transport costs, electric utilities are necessarily distributed widely across the country, which helps to address the concern that firms with different characteristics endogenously select into different regions. Interestingly, not only do our results hold up for this subsample, the economic magnitudes are a bit stronger: our point estimates suggest that an electric utility located in the Deep South has a stock price 8.9% higher than one located in the Middle Atlantic.

Finally, another way to make further progress on the identification problem is to test some of the theory's subsidiary implications. As we demonstrate with the help of a simple model, the RATIO variable should be expected to have the largest impact on the prices of those firms that are least visible outside of their home regions, and whose shares must therefore be absorbed mostly by local investors. Put another way, even though Microsoft is located in the state of Washington, we might not expect its stock price to be too strongly affected by local-market conditions in Washington, since Microsoft is so well-known to investors everywhere else.

Consistent with this hypothesis, we find that the effect of RATIO on stock prices is significantly greater for smaller firms. In particular, for firms in the top quartile of the size distribution (as measured by sales), a move in RATIO from its Middle-Atlantic value to its Deep-South value is associated with only a 4.1% increase in stock prices. By contrast, the corresponding figure for firms outside of the top quartile is more than double, at 9.9%.

In addition to size, we try another proxy for visibility that is more directly connected to the underlying theory: the number of shareholders that a firm has. More precisely, to ensure that we pick up an effect distinct from that of firm size, we measure visibility based on the residual in a regression of the log of the number of shareholders against the log of firm sales. For firms in the top quartile of this visibility proxy, a move in RATIO from its Middle-Atlantic value to its Deep-South value is associated with a 3.2% increase in stock prices. For the remaining, less visible firms, the estimated impact is three times as large, at 9.7%.

The fact that the RATIO variable interacts this way with measures of firm visibility lends further support to our theory, and helps to cut against the alternative that population density—and by extension, RATIO—is just capturing some other unspecified regional factor that matters for stock prices. More precisely, this inference follows as long as we are willing to adopt the identifying assumption that any other unspecified factor does not have a *differential* effect across firms with different degrees of visibility.

The remainder of the paper is organized as follows. In Section 2, we develop a simple model that helps to motivate our tests. In Section 3, we describe the data we use, and the construction of our principal variables. The main hypothesis tests are presented in Section 4. Section 5 discusses related work, and Section 6 concludes.

2. The model

2.1. Basic assumptions

There are N regions of the country. In each region, there are two kinds of firms: "visible" (V) firms and "hometown" (H) firms. In each region, there are also two kinds of investors: "generalists" and "local experts." A generalist can only invest in visible firms, though he is not restricted to those in his region—he can invest in visible firms everywhere. By contrast, a local expert can only invest in the hometown firms in his own region. Thus there are (N+1) completely segmented markets: the N local markets for hometown firms, and the one national market for visible firms.

Denote visible firm i, located in region j, by F_{ij}^{V}. Analogously, denote hometown firm i, located in region j, by F_{ij}^{H}. Firm F_{ij}^{V} has a book value of B_{ij}^{V}, and will pay a liquidating dividend at time 1 of $r_{ij}^{V}B_{ij}^{V}$, where r_{ij}^{V}—which can be loosely thought of as the firm's return on book equity—is a normally distributed random variable with a mean of R_{ij}^{V} and a variance that for simplicity is normalized to one for all firms. Similar notation applies for the H firms.

All investors are assumed to have constant-absolute-risk-aversion (CARA) utility, and the aggregate risk tolerance of investors in region j is given by T_j. A fraction θ of this risk tolerance comes from the generalists, and a fraction $(1-\theta)$ comes from the local experts. The riskless interest rate between time 0 and time 1 is zero.

To simplify the analysis, we further assume that across all firms in all regions, the realizations of the r's are perfectly correlated. This can be thought of as a reduced-form approximation to a one-factor arbitrage-pricing-theory (APT) world, as in Ross (1976), where even within any single region, there are enough hometown firms that a local expert can create a perfectly diversified portfolio that eliminates all idiosyncratic risk.

2.2. Prices of hometown and visible firms

To begin, let us consider the pricing of hometown firms in a given region j. Given the perfect-correlation assumption, we can aggregate these firms into a single combined firm. Call the total time-0 market value of this combined hometown firm V_j^{H}. The expected payoff to this combined firm at time 1 is given by $\Sigma_i(R_{ij}^{H}B_{ij}^{H})$. Similarly, the variance of the payoff to this combined firm is given by $(\Sigma_i B_{ij}^{H})^2$. Since the total risk tolerance for hometown firms in region j is $(1-\theta)T_j$, standard CARA-normal arguments imply that

$$V_j^{H} = \sum_i (R_{ij}^{H}B_{ij}^{H}) - \left(\sum_i B_{ij}^{H}\right)^2 \bigg/ (1-\theta)T_j. \tag{1}$$

By symmetry, it follows that the market-to-book ratio for any one hometown firm i, denoted by Q_{ij}^{H}, is given by

$$Q_{ij}^{H} = R_{ij}^{H} - \left(\sum_i B_{ij}^{H}\right) \bigg/ (1-\theta)T_j. \tag{2}$$

The intuition for Eq. (2) is straightforward. The greater the aggregate book value of firms in a given region, the more risk local investors have to bear, and consequently, the greater is the discount borne by all firms in that region.[2] Moreover, Eq. (2) suggests a very direct empirical test. In particular, we have:

Hypothesis 1. Consider a regression of a hometown firm's market-to-book ratio against: (i) its return on equity (ROE), and (ii) the ratio of the total dollar book value of firms in its region to some proxy for total regional risk tolerance. We expect the former variable to attract a positive coefficient, and the latter to attract a negative coefficient.

The derivation of Eq. (2) is made particularly easy by the assumption that all firms in a region have perfectly correlated returns. As noted above, this is a shortcut way to incorporate the CAPM/APT premise that within a region, investors are fully diversified, and hence only care about systematic risk. But we should emphasize that this sort of CAPM/APT approach is not the only way to generate the result. For example, one might go to the other extreme, and assume that all risk is idiosyncratic, and that each local investor only holds a single hometown stock, instead of a well-diversified local portfolio. If it is also the case that the risk tolerance devoted to any one stock in a region is proportional to its book value—i.e., that a bigger company has a greater probability of getting noticed and hence a larger investor base—a pricing formula identical to that in Eq. (2) once again emerges.

Using the same logic, we can also calculate the market-to-book ratio for any visible firm:

$$Q_{ij}^{V} = R_{ij}^{V} - \left(\sum_j \sum_i B_{ij}^{V}\right) \bigg/ \left(\theta \sum_j T\right). \tag{3}$$

Intuitively, the difference between Eqs. (2) and (3) is that the generalists pool their risk tolerance together across all the regions, but at the same time they have to absorb the book value of all visible firms across all regions.

Thus in addition to Hypothesis 1, we also have:

Hypothesis 2. Consider a regression of a firm's market-to-book ratio against: (i) its ROE; and (ii) the ratio of the total dollar book value of firms in its region to some proxy for total regional risk tolerance. We expect the negative coefficient on the latter variable to be larger in absolute magnitude for hometown firms than for visible firms.

In the context of the model, Hypothesis 2 emerges very starkly, since all visible firms trade in the national market and have exactly the same risk premium. This fact, which is apparent from Eq. (3), implies that for visible firms, the

[2] In the spirit of a segmented-market capital asset pricing model (CAPM) or APT, it is straightforward to extend the model to a setting where individual stocks in a region have different loadings on the local market factor—i.e., different betas. In this case, the price discount for a given stock depends on its own beta, multiplied by a region-level discount factor like that in Eq. (2). However, since we do not test the beta-related implications of the model, we suppress them for simplicity. In other words, we treat all stocks as if they have betas of one.

regression coefficient on the local ratio variable should be exactly zero. In other words, the pricing of visible firms should be completely independent of local-market conditions.

2.3. Taking the model to the data

To implement a test of Hypothesis 1, we create an empirical measure, RATIO, which for each region we define as the aggregate book value of all firms headquartered in the region, divided by the aggregate income of all households living in the region. In using this measure to proxy for the theoretical construct in Eq. (2), several conceptual and practical issues arise.

2.3.1. Which firms go in a region's portfolio?

We use the aggregate book value of all public firms in the numerator of our empirical RATIO variable. This raises two questions. First, Eq. (2) tells us that what should go in the numerator of the theoretically ideal ratio variable is the aggregate book value of *hometown* firms in a region, which we cannot measure. So when we use the book value of *all* firms in a region (both visible and hometown) to build RATIO, this constitutes a form of measurement error, which will bias our estimates downwards. This sort of measurement error will be less of a problem to the extent that $\Sigma_i B_{ij}^{\mathrm{H}}$ and $\Sigma_i B_{ij}^{\mathrm{V}}$ are highly correlated at the regional level.

Second, our empirical measure does not include the book value of *private* firms in the region. There are obvious data-related impediments to incorporating private firms, but we would also argue that leaving them out is the conceptually more sensible thing to do. This is because the evidence suggests that the market for privately-held firms is largely segmented from that for public firms (Moskowitz and Vissing-Jorgensen, 2002). Intuitively, if most investors in public equity are not also private-firm owners, then the stock of private equity in a region is not relevant for the pricing of public equity. If so, it is appropriate to compute RATIO as we do, excluding the book value of private firms in a region.

2.3.2. Is regional income a good proxy for regional risk tolerance?

When we use aggregate regional income in the denominator of our RATIO measure, we are implicitly assuming that each investor's risk tolerance is proportional to his labor income, which implies that regional risk tolerance is proportional to total regional labor income. Since individual risk tolerance is just an exogenous parameter in the CARA model, there is no fundamental logical problem in assuming that this parameter covaries with income in the population. Nevertheless, it would clearly be more satisfying to have a model in which something like this feature emerges endogenously.

To this end, we have also experimented with a constant-relative-risk-aversion (CRRA) version of the model. Unfortunately, solving for the equilibrium price in closed form in a CRRA setting with outside labor income is not possible. We are, however, able to establish the following partial-equilibrium result: as an investor's outside labor income increases, so does his allocation to the risky asset, holding fixed his CRRA risk-aversion coefficient. Thus an increase in labor income acts like an increase in risk tolerance, as we effectively assume in our implementation of the CARA model. Unfortunately, even the derivation of this very simple result gets quite messy.

Thus for the sake of expositional clarity, we have chosen to retain the simpler CARA model, while explicitly acknowledging the leap that we are making when we proxy for risk tolerance in a region with total income. We should stress, however, that we are *not* assuming that individual risk tolerance varies systematically across regions. Rather, we are assuming that, within any region, there is a correlation between an individual's labor income and his risk-tolerance parameter. The magnitude of this correlation can be kept fixed across all regions.

2.3.3. Is book value a sensible measure of scale?

Our model assumes that the dollar variance of a firm's cashflows is related to the firm's book value. Although this would seem to be a reasonable assumption, one can think of other plausible scaling variables. In this spirit, we also try scaling by firm cashflow, instead of firm book value, which corresponds to the assumption that the dollar variance of a firm's cashflows is related to the current level of its cashflow. This involves two changes in our methodology: (i) when we compute RATIO, we now use the total cashflow of all firms headquartered in a region as the numerator; and (ii) in parallel, we use a firm's ratio of price-to-cashflow, rather than its market-to-book, as the valuation measure on the left-hand-side of our regressions.

One important virtue of this alternative approach to scaling is that it allows us to address the following concern. It is likely that a firm located in the Middle Atlantic will have to pay higher prices for physical assets such as land and buildings than an otherwise similar firm located in the Deep South. This will tend to push up the measured value of RATIO for Middle Atlantic firms, and push down their market-to-book values, potentially inducing a mechanical negative correlation between these two variables. When we rescale everything by cashflows instead of book values, any influence of the book values of land and buildings is removed. As it turns out, the rescaling leads to very similar results, which is reassuring on this score.

2.3.4. How big is a region?

The model provides no guidance as to the level of regional aggregation (city, state, Census region) at which we should ideally measure RATIO. In order to think about this issue in an a priori, evidence-based fashion, we begin with the following exercise, using Barber and Odean's (2000) data on individual-investor holdings from a discount brokerage firm. Taking a December 1995 snapshot from this data set, and restricting attention to the 3,000 largest U.S. stocks, we run a multivariate regression in which the dependent variable is a dummy that equals one if an individual investor owns a given stock, and in which the independent variables are three dummies that equal one if the investor and the firm are in the same Metropolitan Statistical Area (MSA), the same state, and the same Census region, respectively.

The coefficients on these three dummies are shown in Panel A of Table 1. They imply that if the investor is in the same MSA as the firm, the probability of owning the stock goes up by 1.41 percentage points (this is the sum of the coefficients on all three dummies). If the investor is in the same state as the firm, but not the same MSA, the probability goes up by 0.47 percentage points (this is the sum of the coefficients on the state and Census region dummies). If the investor is in the same Census region as the firm, but not the same state, the probability goes up by 0.18 percentage points. All of these numbers are strongly statistically significant. They are also economically significant, given that the baseline probability of owning a completely non-local stock (the constant term in the regression) is only 0.33%.

On the one hand, these results make it clear that the closer an investor is to a stock, the stronger is the degree of local bias. This is an intuitive outcome. At the same time, they also show that substantial local bias exists at the Census region level, and crucially, that this bias is present even for investors who are not in the same MSA or state as the firm. The question then becomes: given this configuration, what is the most informative level of aggregation for the purposes of measuring the RATIO variable?

In our view, the conceptually most sensible basis for making a choice is to think in terms of the *population-weighted excess demand* generated by local bias. To see the logic of this approach, consider an extreme case. Suppose we knew that anybody who lived within a mile of a firm's headquarters was enormously locally biased, and had a 100% probability of holding the firm's stock. Would we want to measure the RATIO variable for only the small number of households this close to each firm, on the notion that local bias is strongest at this very fine neighborhood level of aggregation? Clearly not. This is because the total dollar excess demand that these few households contribute is very small, so measuring their income cannot tell us much about the relevant factor for the stock price, which is the *total incremental demand for the stock that is induced by local bias*.

To develop the implications of this idea, Panel B of Table 1 presents some illustrative calculations for the case of Indianapolis, which is a roughly median-sized MSA. The population of Indianapolis is 1.745 M in 2005. The population of Indiana outside of Indianapolis is 4.521 M. And the population of the Midwest Census region outside of Indiana is 39.865 M. Using these figures, and the local-bias estimates above, Panel B shows that of the total population-weighted excess demand for a hypothetical Indianapolis-based stock that is generated by local bias, 21% comes from investors in the Indianapolis MSA, 18% comes from investors living in Indiana but outside of Indianapolis, and 61% comes from investors living in the Midwest Census region but outside of Indiana.

These results provide a strong ex ante basis for measuring RATIO at the Census-region level, as we do in our baseline specifications below. This logic holds even though, for any single investor, local bias is stronger at the MSA level; this intensity-of-bias effect is overwhelmed by the fact that there are far fewer locally-biased investors in an MSA than in a Census region. In the case of Indianapolis, if we were to measure RATIO at the MSA level, we would be ignoring 79% of the total population-weighted excess demand that is associated with local bias.

2.3.5. *Is there local bias in all regions of the country?*

The brokerage-account data also allow us to verify a key premise of our model—namely, that investors in all regions of the country exhibit similar degrees of local bias. In particular, we want to make sure that this bias is not substantially weaker in regions with relatively few local firms, like the Deep South. If Deep-South investors were to respond to the scarcity of locally-available stocks by investing more broadly outside their home region than,

Table 1

Local bias of individual investors at the MSA, state and census-region levels

In Panel A, we run a multivariate regression in which the dependent variable is a dummy that equals one if an individual investor owns a given stock, and in which the independent variables are dummies that equal one if the investor and the firm are in the same MSA, same state, and same Census region, respectively. The regression is based on brokerage-account data on the holdings of investors as of December 1995, and is restricted to the 3,000 largest U.S. stocks. Standard errors are in parentheses; they are clustered by MSA. In Panel B, we use the point estimates from Panel A, along with population information as of 2005 to compute, for a hypothetical Indianapolis-based firm, the share of total Midwest-region population-weighted local bias that arises at the level of the MSA, the state, and the Census region. The "local bias" entries in Row 1 of Panel B are based on the regression coefficients in Panel A.

Panel A: Regression of stock ownership on MSA, state and census region indicators

	Constant (%)	Same MSA (%)	Same state (%)	Same census region (%)
Probability of owning stock	.33 (.01)	.94 (.33)	.29 (.07)	.18 (.03)

Panel B: Contribution of MSA, state and region to population-weighted local bias: Indianapolis example

	Indianapolis MSA (%)	Indiana ex Indianapolis MSA (%)	Midwest region ex Indiana (%)
1. Local bias	1.41	.47	.18
2. Share of region population	3.78	9.80	86.42
3. Pop-weighted bias (1×2)	.053	.046	.156
4. Share of region pop-weighted bias	20.78	18.04	61.18

Table 2
Local bias of individual investors by census region

The entries are measures of the local bias of individual investors, using brokerage-account data on the holdings of investors as of December 1995. The first column shows the mean probability that an investor holds one of the 3,000 largest U.S. stocks. The second column reports the coefficient in a regression in which the dependent variable is a dummy that equals one if an individual investor owns a given stock, and in which the independent variable is a dummy that equals one if the investor and the firm are in the same Census region. Standard errors are in parentheses; they are clustered by Census region. The final column reports the size of the increase in probability of holding a stock if it is local (column 2) relative to the mean probability of holding a stock (column 1). The first row shows the calculations for all investors in the U.S. The remaining rows include subsamples of investors located in each of the nine Census regions.

	Mean probability stock held by investor (%)	Increase in probability if local stock (%)	Increase relative to mean probability (%)
All investors	.51	.59 (.07)	116
New England	.48	.41 (.06)	85
Middle Atlantic	.53	.48 (.04)	91
Midwest	.53	.41 (.05)	77
Plains	.45	.79 (.03)	176
Atlantic Coast	.54	.25 (.04)	46
Deep South	.46	1.12 (.04)	243
Southern Plains	.50	.90 (.03)	180
Mountain	.49	.87 (.05)	178
West Coast	.51	.76 (.04)	149

say, Middle-Atlantic investors, this would tend to mitigate the pricing effects identified in our model.

The first row of Table 2 notes that, across the entire sample, the mean probability that a given investor holds one of the 3,000 largest U.S. stocks in December 1995 is 0.51%. Moreover, this probability goes up by 0.59 percentage points (an increase of 116% relative to the unconditional value) if the stock is located in the same Census region as the investor. This incremental effect is based on a regression of a dummy variable for an investor holding a given stock against a dummy for the investor and the firm being located in the same Census region.

In the next nine rows of Table 2, we perform the same analysis separately for investors living in each of the nine different Census regions. As can be seen, there is significant local bias in each of the regions taken individually. Moreover, this local bias is actually most pronounced for investors in the Deep South, where the probability of an investor owning a stock goes up by 1.12 percentage points (an increase of 243% relative to the mean value) if the stock is local. Local bias is also very strong in other low-RATIO regions, such as the Mountain region and the Plains. Thus it appears that investors in regions with relatively few local stocks to choose from *do not* adapt to this scarcity by becoming less locally-oriented. We conclude from this exercise that the underlying premise of our model is on firm ground.

3. Data

3.1. Sources

Our data on personal income and other regional demographic variables come from a database produced by the Bureau of Economic Analysis (BEA), the Personal Income and Population Summary Estimates. It is available on the BEA's Web site, www.bea.doc.gov, going back to 1969. We limit our analysis to the time period from 1970 through 2005. Our data on firms come from the Center for Research in Security Prices (CRSP) and Compustat. From CRSP, we obtain stock prices and shares outstanding for NYSE, Amex and Nasdaq stocks. From Compustat, we obtain annual information on a variety of accounting variables, as well as the locations of firms' headquarters. To be included in our sample, a firm must first have the requisite financial data on CRSP and Compustat, and must have headquarters in the lower 48 states or in the District of Columbia (i.e., we drop firms located in Alaska and Hawaii).

When we run our regressions, we exclude observations on firms with book equity values of less than ten million dollars, as well as those with one-digit SIC codes of 6, which are in the financial-services industry. However, the book values (cashflows) of these firms are kept for the purposes of computing the aggregate book value (cashflow) of firms in a region, which is a key component of our RATIO variable. It should also be noted that our results are if anything slightly stronger if we include the observations on the smallest firms.

3.2. Variable definitions

The market equity value of a firm (M), defined as the combined value of all common stock classes outstanding, is taken from CRSP as of fiscal year end. For the book equity value (B), we use Compustat data item 60. Our primary dependent variable is the log of the ratio of market equity to book equity, i.e., $\log(M/B)$. We take logs because the raw market-to-book ratio is highly skewed, and the log transformation results in a variable that is much closer to being symmetrically distributed. However, we obtain similar (albeit somewhat less precisely estimated) results if we instead work with the raw market-to-book ratio. Alternatively, we can use the book-to-market ratio as the dependent variable, which also leads to results very similar to those we report below, though of course with all of the signs reversed. Finally, as noted above, we also experiment with an entirely different valuation measure: the log of a firm's price-to-cashflow ratio, $\log(M/C)$, where cashflow C is net income (item 172) plus depreciation (item 14). In calculating this last ratio, we exclude firms with negative values of cashflow.

The BEA database reports total personal income by state and breaks the personal income down by its various parts, including dividend income. Our main independent variable of interest, RATIO, is the ratio of total book equity in a region to total personal income in that region. In calculating personal income, we exclude dividend income, on the notion that keeping it in might induce an artificial, hard-wired relationship between RATIO and stock prices.[3] We calculate RATIO both for Census regions (nine regions in all) and states, though we focus primarily on the Census-region results. The BEA database also reports per capita income and population density by state, which we can aggregate up to get analogs for Census regions.

A firm's return on book equity (ROE) is its net income (Compustat item 172) divided by its previous-year book equity (item 60 lagged 1 year). R&D expenditures and sales are items 46 and 12, respectively, and we use these to create an R&D-to-sales ratio. The log market-to-book ratio, ROE, and R&D-to-sales ratio are all winsorized at the 1% and 99% levels. When the R&D variable is missing, we set its value to zero; however, we also include in all our regressions a dummy that equals one when a firm does not report R&D.

For the purposes of one of our robustness checks, we create a dummy variable that equals one for a conglomerate, which we define as a firm that operates in more than one business segment. Information regarding firm segments on Compustat only begins in 1983. So our analysis involving this variable is limited to the subsample that runs from 1983 to 2005. If a firm is missing segment data, we assume that it is not a conglomerate, and set the dummy variable to zero. In another robustness check, we drop observations corresponding to any firm that belongs to a dominant industry in its region. For each region and each year, we calculate the book value of firms in each two-digit SIC industry, and we deem an industry to be dominant if it accounts for more than 10% of the total book value in that region. Note that we only drop these dominant-industry firms from the left-hand-side of our regressions, but keep them in when calculating the total book equity of firms in a Census region or state.

3.3. *Anatomy of the RATIO variable*

Table 3 provides some summary statistics for the RATIO variable. In Panel A, we display the value of RATIO for each Census region once every 5 years between 1970 and 2005, along with both cross-sectional and time-series means and standard deviations. As can be seen, the Middle Atlantic region has consistently had the highest values of RATIO, averaging 0.77 over the sample period. New England comes in second, with an average RATIO of 0.60. At the other extreme, the Deep South has the lowest average value over the entire sample period, at 0.21. As an inspection of the table suggests, these differences across Census regions are highly persistent: if we rank regions by their values of RATIO, and regress the ranks on their once-lagged values, we get a coefficient of 0.98.[4]

In Panel B, we show the analogous data at the state level. In keeping with the patterns seen in Panel A, high-population-density states like Connecticut, New York, and Illinois are among those with the highest average values of RATIO, while low-density states like Wyoming, Montana, West Virginia, and Vermont rank near the bottom. Thus both panels of the table make it clear that there is a close link between the RATIO variable and population density.

At the same time, the state data are more volatile in the time series dimension than the Census-region data, and yield some striking anomalies. For example, the state of Arkansas, which has very low values of RATIO in the 1970s and 1980s, is by 2005 the fourth-highest-RATIO state in the country, with a value of 1.17. This is almost entirely due to the fact that Wal-Mart is headquartered in Arkansas; without Wal-Mart, Arkansas' RATIO in 2005 would be only 0.39. Even more extreme in recent years is the effect of Berkshire-Hathaway on Nebraska, which has the single highest value of RATIO in 2005, at 2.33; coincidentally, Nebraska would also have a RATIO of 0.39 in 2005 without Berkshire-Hathaway. Evidently, in small states, the observed value of RATIO can be extremely sensitive to the presence of a single large firm.

Also noteworthy is Delaware, which, over the 1970–2005 sample period, has the highest average value of RATIO among all states, at 1.54. This likely reflects Delaware's dominant role in the market for incorporations, which might lead a disproportionate number of firms to also list their nominal headquarters as being in Delaware, even if they do not have much of an operating presence in the state.[5] From our perspective, this can be thought of as a form of measurement error in RATIO, to the extent that local bias among investors is primarily a function of operating presence. The measurement-error point is reinforced by the fact that Delaware's RATIO falls dramatically, to 0.40, by 2005. This drop is largely due to MBNA Corp.'s merger with Bank of America, which causes MBNA's assets to "disappear" when we compute Delaware's RATIO, despite there being no change in where MBNA actually does business.

These observations provide a second motivation—beyond that discussed in Section 2.3.4 above—for using Census-region measures of RATIO in our baseline specifications. A further advantage of aggregating up to the Census-region level is that we can hope to smooth out some of the noise in RATIO that arises at the state level. However, we do return to the state-level measures in several of our robustness checks.

[3] If, controlling for ROE, higher dividends are associated with higher stock prices, it is conceivable that a region with a lot of dividend income—and hence a lower value of RATIO, if dividends are included in the calculation of RATIO—would show up as having higher valuations on average.

[4] Alternatively, if we regress RATIO on both Census-region and year effects, the partial *R*-squared of the region effects is 0.82, again suggesting that most of the variation in RATIO is across Census regions, rather than over time.

[5] See Bebchuk and Cohen (2003), who report that as of 1999, 58% of all public firms were incorporated in Delaware.

Table 3
Summary statistics for RATIO, 1970–2005

The entries are values of RATIO, the ratio of total book equity to total personal income (less dividends), in a given region. Panel A reports RATIO for the nine Census regions in every fifth year, starting in 1970, as well as the time-series means and standard deviations of RATIO (using all the years, 1970–2005). In addition, for each of the years shown, the cross-sectional mean and standard deviation of RATIO are reported. Panel B is similar, except that RATIO is calculated for states instead of Census regions.

	1970	1975	1980	1985	1990	1995	2000	2005	Mean	S.D.
Panel A: Census regions										
New England	.51	.61	.64	.54	.51	.61	.66	.71	.60	.06
Middle Atlantic	.75	.83	.85	.69	.59	.66	.90	1.03	.77	.13
Midwest	.54	.57	.56	.58	.52	.54	.50	.52	.54	.03
Plains	.23	.29	.32	.32	.31	.42	.53	.62	.40	.11
Atlantic Coast	.32	.35	.32	.33	.31	.36	.48	.48	.38	.06
Deep South	.08	.14	.15	.18	.15	.24	.26	.25	.21	.05
Southern Plains	.65	.67	.65	.51	.46	.51	.70	.73	.59	.09
Mountain	.18	.23	.19	.24	.20	.25	.40	.29	.26	.06
West Coast	.29	.33	.33	.32	.30	.39	.63	.60	.42	.12
X-sectional mean	.40	.45	.45	.41	.37	.44	.56	.58		
X-sectional S.D.	.23	.23	.24	.17	.15	.15	.19	.24		
Panel B: States										
New England										
Connecticut	1.24	1.42	1.54	1.13	1.07	1.15	1.08	1.48	1.26	.18
Massachusetts	.24	.33	.31	.37	.32	.48	.65	.48	.44	.10
Maine	.10	.11	.11	.14	.12	.16	.15	.29	.14	.04
New Hampshire	.12	.16	.19	.15	.13	.16	.07	.13	.13	.04
Rhode Island	.28	.33	.37	.29	.34	.38	.42	.48	.37	.07
Vermont	–	.01	.01	.02	.04	.10	.09	.06	.07	.04
Middle Atlantic										
New Jersey	1.24	1.18	1.14	.62	.51	.60	.98	.62	.79	.28
New York	.69	.81	.87	.86	.74	.83	1.15	1.55	.92	.22
Pennsylvania	.52	.60	.58	.42	.36	.39	.38	.45	.46	.09
Midwest										
Illinois	.76	.81	.87	.90	.82	.84	.72	.71	.82	.06
Indiana	.15	.18	.18	.17	.18	.20	.21	.41	.21	.05
Michigan	.70	.65	.53	.60	.62	.51	.43	.34	.54	.12
Ohio	.45	.50	.54	.54	.38	.47	.55	.57	.50	.06
Wisconsin	.15	.24	.22	.21	.22	.29	.24	.33	.24	.04
Plains										
Iowa	.05	.08	.09	.12	.15	.18	.13	.22	.14	.05
Kansas	.06	.10	.10	.10	.07	.10	.12	.10	.10	.02
Minnesota	.38	.50	.52	.50	.42	.52	.63	.89	.58	.13
Missouri	.30	.39	.42	.41	.38	.47	.45	.37	.41	.04
North Dakota	–	–	.00	.02	.00	.02	.07	.00	.03	.04
Nebraska	.15	.25	.27	.37	.51	.98	2.05	2.33	1.11	.78
South Dakota	–	.03	.06	.02	.07	.12	.14	.07	.07	.05

Table 3 (*continued*)

	1970	1975	1980	1985	1990	1995	2000	2005	Mean	S.D.
Atlantic Coast										
Delaware	1.72	2.05	1.59	1.83	1.67	1.04	1.30	.40	1.54	.49
District of Columbia	.25	.41	.54	.58	.81	1.65	1.45	.60	.88	.47
Florida	.19	.23	.19	.17	.11	.16	.19	.19	.18	.03
Georgia	.26	.34	.31	.47	.47	.50	.67	.68	.51	.13
Maryland	.22	.25	.23	.22	.19	.22	.25	.25	.22	.02
North Carolina	.30	.34	.32	.31	.28	.43	.72	1.06	.46	.22
South Carolina	.11	.14	.12	.12	.12	.15	.15	.10	.13	.02
Virginia	.60	.59	.59	.52	.52	.54	.75	.71	.60	.09
West Virginia	.00	.02	.00	.00	.04	.05	.04	.05	.03	.02
Deep South										
Alabama	.03	.09	.10	.14	.15	.21	.32	.30	.20	.08
Kentucky	.10	.16	.15	.16	.15	.17	.12	.18	.15	.02
Mississippi	–	.09	.12	.17	.17	.24	.20	.14	.18	.07
Tennessee	.07	.18	.18	.19	.13	.30	.32	.28	.25	.08
Southern Plains										
Arkansas	.05	.08	.11	.19	.36	.62	.89	1.17	.45	.35
Louisiana	.07	.09	.08	.08	.08	.10	.13	.16	.10	.03
Oklahoma	.27	.30	.29	.17	.12	.18	.29	.41	.23	.08
Texas	.96	.98	.93	.71	.61	.63	.83	.81	.77	.13
Mountain										
Arizona	.22	.26	.20	.20	.13	.16	.17	.19	.19	.04
Colorado	.28	.39	.28	.38	.36	.44	.88	.51	.46	.17
Idaho	.10	.16	.17	.22	.23	.25	.46	.40	.25	.11
Montana	–	–	.03	.05	.00	.10	.10	.04	.06	.04
New Mexico	–	.02	.03	.02	.05	.06	.08	.02	.06	.04
Nevada	.06	.11	.10	.16	.15	.30	.27	.32	.22	.08
Utah	.19	.24	.22	.22	.17	.21	.12	.17	.19	.04
Wyoming	–	–	.02	–	.00	.00	–	.00	.01	.01
West Coast										
California	.34	.36	.37	.34	.31	.41	.63	.63	.44	.11
Oregon	.07	.17	.20	.21	.26	.30	.21	.13	.22	.05
Washington	.11	.19	.20	.20	.23	.27	.79	.70	.40	.24
Cross-sectional mean	.30	.35	.34	.33	.31	.38	.48	.46		
Cross-sectional S.D.	.37	.40	.36	.33	.31	.33	.43	.45		

Table 4
RATIO and its components vs. population density, census regions, 1970–2005

Each entry is the time-series mean of cross-sectional regression coefficients, estimated each year from 1970 to 2005. The dependent variables include: (i) the log of RATIO, the ratio of total book equity to total personal income (less dividends) in a given Census region; (ii) the log of total book equity per capita in a Census region; and (iii) the log of per capita income in a Census region. Each of these variables is regressed one at a time against population density in the region.

Dependent variable	Log(RATIO)	Log(Region book/capita)	Log(Region income/capita)
Coefficient on pop. density×100	.29 (.05)	.36 (.04)	.07 (.01)
R-squared	.42	.48	.35

In order to better understand what drives the RATIO variable, we take logs and write the log of RATIO as the log of regional book value per capita, minus the log of regional income per capita. Using this decomposition, we can, in any cross-section, ask how much of the variance of the log of RATIO is coming from each of these two terms. The answer is that the lion's share comes from the log of book value per capita: at the Census-region level, an average of 69% of the total variance of the log of RATIO comes from the log of book value per capita.

Table 4 relates this decomposition explicitly to regional differences in population density. Focusing again on Census regions, we run annual cross-sectional regressions with three different dependent variables: (i) the log of RATIO; (ii) the log of book value per capita; and (iii) the log of income per capita. In each case, the sole explanatory variable is regional population density. In Column 1, it can be seen that the log of RATIO is highly correlated with population density; this confirms the informal impressions from Table 3. The *R*-squared in the univariate regression averages 0.42, which is equivalent to a correlation coefficient of 0.65. Columns 2 and 3 demonstrate that both components of the log of RATIO are also positively correlated with population density, but that the book value term is considerably more sensitive to population density than the income term, with an average regression coefficient of 0.36 vs. 0.07. In other words, as population density goes up, both book value per capita and income per capita also rise, but the former effect is much stronger than the latter, so that on net, RATIO increases as well.

With a little reflection, the strong effect of population density on book value per capita—and, ultimately, on RATIO—makes intuitive sense. There are many reasons why firms would prefer, all else equal, to locate their headquarters and/or their major operating facilities in densely populated areas: better infrastructure (e.g., large international airports); access to a deeper and higher-quality labor pool; etc. Indeed, these sorts of agglomeration effects provide perhaps the most natural way of thinking about the root source of variation in our RATIO measure.

4. Empirical results

4.1. Baseline specification

Our baseline specification is designed to test Hypothesis 1. The dependent variable is the log of the market-to-book ratio for a firm. The independent variables are RATIO, firm ROE, firm R&D-to-sales, a dummy for whether the firm reports R&D expenditures, a set of four-digit SIC industry dummies, and dummies for exchange listing (NYSE, Amex or Nasdaq).[6] To further protect against hardwiring, we recalculate the RATIO variable for each firm-year observation so that the numerator of RATIO omits the book value of the firm in question. That is, for an observation on firm *i* in year *t*, the corresponding RATIO variable includes the book value of all the *other* firms in *i*'s region in year *t*, but does not include *i*'s book value. This ensures that when we obtain a negative coefficient in a regression of log market-to-book against RATIO, it is not coming simply because the same book value is in the denominator of the left-hand-side variable and the numerator of the right-hand-side variable. It also implies that while the values of RATIO for firms in a given region/year cell are very highly correlated, they are not literally identical. As it turns out, however, this adjustment is inconsequential. We obtain essentially identical results with a naïve approach that makes no such correction.

4.2. Statistical inference

Given the nature of our data, we need to think carefully about the correlation structure of the residuals, and about the resulting implications for how we calculate standard errors. First, note that in any given year, we only have nine effectively independent observations on RATIO if we are working at the Census-region level, and 49 if we are working at the state level. In other words, we expect a high degree of cross-correlation in the residuals at a given point in time.

As one method for dealing with this cross-correlation, we take a Fama-MacBeth (1973) approach, running a separate cross-sectional regression each year from 1970 to 2005—a total of 36 regressions in all. We then compute the means of the annual regression coefficients. Finally, we evaluate the significance of the means based on an in-sample estimate of the time-series variance of the annual coefficients, one that adjusts for serial correlation in these coefficients.

However, as Petersen (2005) points out, the Fama-MacBeth (1973) approach—even with a serial-correlation adjustment—can lead to understated standard errors if

[6] We include a dummy for R&D expenditures because firms involved in R&D may have significantly different market-to-book ratios than those not involved in R&D (see, e.g., Chan, Lakonishok, and Sougianis, 1999).

there are fixed or slowly-decaying effects in the data. To see this point most clearly, consider an extreme example where the data for each of the 36 years in our sample are literally identical. In this case, the annual regression coefficients will also be identical, so the Fama-MacBeth method will inappropriately generate estimated standard errors that approach zero.

As an alternative that does better in the presence of fixed or slowly-decaying effects, Petersen (2005) suggests the use of a single pooled regression with clustered standard errors. To implement this procedure, we pool all the data, add year dummies, and also allow each of the control variables other than RATIO to take on a different value each year (i.e., these controls are all interacted with year dummies). We then cluster the standard errors at the region level.

To check that this approach is robust in the presence of fixed effects in our setting, we try the following experiment. First, we take a single year at random from our data set, and run the regression for just that year in isolation. Then, we create a fake 36-year panel which consists of 36 repetitions of that same 1 year, and apply our pooled-plus-clustering specification to the fake panel. Ideally, and in contrast to the Fama-MacBeth (1973) procedure, we should get the same standard errors with the 1-year regression as with the fake 36-year panel, because there is no more information in the latter—i.e., the method should not be fooled by the presence of more observations if these observations are purely redundant. And indeed, this is what happens.[7] We conclude that Petersen's (2005) arguments carry over to our setting, and that the pooled-plus-clustering methodology is robust to even the most extreme kind of fixed effects.

As a final reality check on our standard errors, we also try a "collapsed" version of our baseline specification. To implement this, we first run every year a cross-sectional regression of log market-to-book against all of our controls *except* RATIO, and use this regression to generate firm-level residual values of log market-to-book. Next, we collapse these firm-level residuals at the region-year level—i.e., we compute an equal-weighted average of the firm-level residuals within each region-year cell. In the case of Census regions, this gives us nine observations on residual log market-to-book each year—one for each region—and a total of 324 observations for the entire 36-year sample period ($324 = 9 \times 36$). We then run a simple panel regression on these 324 observations, with RATIO and year dummies as the only right-hand side variables, again clustering the standard errors at the region level to account for serial correlation in the residuals.

This approach is appealing because it makes transparently clear that, in computing the standard errors, we are not claiming to have anything more than nine independent observations per period. And comfortingly, it leads to results that are very close to those from the pooled regressions. Moreover, a modification of the collapsing technique also allows us to provide a simple graphical illustration of our results, which we will turn to shortly.

4.3. Baseline results

Table 5 gives a detailed overview of our baseline results. RATIO is measured at the Census-region level. When we use the Fama-MacBeth (1973) approach, the coefficient on RATIO takes on the predicted negative sign in 35 of the 36 regressions. Across all of the regressions, the mean value of the coefficient is −0.150, with a Fama-MacBeth (serial-correlation adjusted) standard error of 0.020. Not surprisingly, the coefficients on ROE and R&D-to-sales are both positive in each of the 36 regressions.

As noted above, there is a concern that in our setting the Fama-MacBeth (1973) approach may produce understated standard errors. The results from the pooled and collapsed regressions bear out this concern. Although the point estimates for the coefficient on RATIO are qualitatively similar to those from the Fama-MacBeth specification, at −0.136 and −0.105 respectively, the standard errors are larger, at 0.034 and 0.029. Nevertheless, even with these more conservative standard errors, the coefficient on RATIO remains statistically significant at the 1% level. In particular, the *t*-statistic from the pooled regression is 4.00.

In all the variations that follow, we use the pooled-plus-clustering approach as our primary method of inference. However, the collapsing methodology is especially useful for providing some further intuition as to why, in spite of all the concerns about both cross-correlation and time-series correlation, we are able to obtain statistically significant estimates for the coefficient on RATIO. In Fig. 1, we take the collapsing approach a step further, by averaging the data over time. We plot the time-averaged value of residual log market-to-book for each Census region against the time-averaged value of RATIO, where this averaging is done over the entire sample period. This means that we have boiled all the data down to just nine observations—discarding, among other things, any potentially useful information that might be embodied in year-to-year variation in RATIO. Yet, as can be seen from the scatterplot, these nine observations alone tell a pretty clear story: a regression based on the nine data points in the figure yields a coefficient on RATIO of −0.120, a standard error of 0.045, and a *t*-statistic of 2.67. Thus it should not be too surprising that no matter how stringently we adjust our standard errors to account for various forms of correlation in the data—i.e., no matter how much we discount the effective number of independent observations—we still obtain statistically significant results.

To get a sense of economic magnitudes, recall that the average value of RATIO in the Middle Atlantic is 0.77, while the average value in the Deep South is 0.21. Thus if a firm moves from the Middle Atlantic to the Deep South, holding all else equal, the implied increase in the log of

[7] For example, if we run the regression for just the single year 1995, we get a point estimate for the coefficient on RATIO of −0.264, with a standard error of 0.138. If we run the regression for a fake panel with 36 repetitions of the 1995 data, we get an identical point estimate, and a little-changed standard error of 0.131. Similar results obtain if we start with any other year in our sample period.

Table 5
RATIO and stock prices, detailed Census-region results

The dependent variable is the log of the ratio of market equity to book equity for a company. The independent variables are RATIO, the ratio of total book equity to total personal income of the Census region in which the company is located, along with the company's ratio of R&D to sales, and return on equity (ROE). Also included in the regressions (but not shown) are a dummy variable which equals one if the company does not report R&D expenditures, a set of four-digit SIC industry dummies, and dummies for exchange listing (NYSE, Amex or Nasdaq). Entries are the coefficients for RATIO, R&D to sales, and ROE, and the *R*-squared of the cross-sectional regressions by year. Also reported are the time-series means of these yearly coefficients, the Fama-MacBeth (1973) serial-correlation-adjusted standard errors, and the fraction of the years in which the regression coefficients have the predicted signs. The last two lines report the results from: (i) a single pooled regression in which the standard errors are clustered at the region level; (ii) a collapsed regression in which there is only one averaged observation on residual log market-to-book for each region-year cell, and in which the standard errors are again clustered at the region level. In the pooled regression, there are year dummies, and the coefficients on all control variables other than RATIO are allowed to vary by year. In the collapsed regression, there are only year dummies in addition to the RATIO variable. Statistical significance at the 1% level indicated by ***.

	RATIO	R&D to sales	ROE	*R*-squared
1970	−.077	3.85	3.29	.66
1971	−.195	6.58	2.94	.62
1972	−.141	6.14	4.05	.66
1973	−.184	9.13	3.74	.61
1974	−.139	6.89	2.10	.54
1975	−.122	7.66	2.72	.58
1976	−.125	7.28	2.29	.56
1977	−.109	4.38	2.36	.57
1978	−.110	4.54	2.35	.58
1979	−.086	5.68	1.81	.58
1980	−.127	4.67	2.20	.64
1981	−.133	2.49	1.88	.54
1982	−.176	.813	1.47	.52
1983	−.112	.890	1.18	.49
1984	−.108	.469	1.21	.46
1985	−.120	.546	.915	.46
1986	−.067	.874	1.14	.47
1987	−.224	.320	1.02	.42
1988	−.262	.163	.975	.38
1989	−.310	.202	1.10	.43
1990	−.333	.457	1.30	.47
1991	−.140	.663	1.29	.48
1992	−.190	.257	.864	.39
1993	.054	.255	.585	.33
1994	−.218	.213	.667	.35
1995	−.264	.349	.709	.40
1996	−.268	.380	.564	.38
1997	−.162	.260	.475	.34
1998	−.251	.295	.580	.38
1999	−.282	.393	.285	.46
2000	−.091	.364	.663	.42
2001	−.054	.106	.449	.39
2002	−.039	.088	.690	.37
2003	−.073	.174	.425	.36
2004	−.023	.238	.403	.31
2005	−.135	.187	.560	.29
Avg. coefficient	−.150***	2.08	1.79	
F-M std. error	(.020)	(2.73)	(1.19)	
# With predicted sign	35/36	36/36	36/36	
Pooled regression	−.136*** (.034)			
Collapsed regression	−.105*** (.029)			

market-to-book based on the pooled estimate is 0.076 ($0.136 \times (0.77-0.21) = 0.076$), i.e., the firm's stock price goes up by about 7.9% ($\exp(0.076)-1 = 0.079$).

4.4. Additional controls

In Table 6, we add a variety of further controls to our baseline specification. For compactness, we now display in each row of the table only the summary estimates associated with the pooled regression model. In Row 1, we reproduce our baseline result from the pooled specification—a coefficient on the RATIO variable of −0.136. In Row 2, we add to the regression region per-capita income. This variable is itself completely insignificant, while the coefficient on RATIO is little changed, at −0.127. Intuitively, this tells us that our results for RATIO are driven almost entirely by variation in the numerator

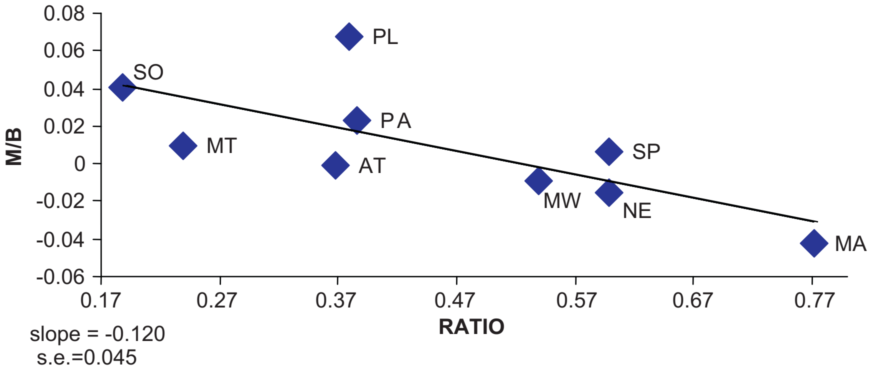

Fig. 1. Scatterplot of residual log market-to-book vs. RATIO. The figure plots averages of residual log market-to-book against time-averaged values of RATIO at the Census-region level. The residuals are based on cross-sectional regressions of log market-to-book against all the control variables in Table 5, excluding RATIO. These residuals are then averaged both across firms in a Census region, and then over the entire 1970–2005 sample period. Census regions are indicated as follows on the plots: New England is NE; Middle Atlantic is MA; Midwest is MW; Plains is PL; Atlantic Coast is AT; Deep South is SO; Southern Plains is SP; Mountain is MT; and Pacific is PA. We also display the coefficient from the nine-data-point regression of average residual log market-to-book against average RATIO, along with the associated standard error.

Table 6
RATIO and stock prices, additional controls

This table presents the results of variations on the pooled regression presented in Table 5. The dependent variable is the log of the ratio of market equity to book equity for a company. In addition to the independent variables in Table 5, Row 2 adds region per capita income. Row 3 adds region population density. Row 4 adds the growth rate of region income from year t+1 to t+3. Row 5 adds the average firm ROE over years t+1 through t+3. Row 6 adds the growth rate of firm sales from year t+1 to t+3. Row 7 adds all the future controls in Rows 4–6 simultaneously. Row 8 removes observations corresponding to industries that account for more than 10% of total book value in a region. Row 9 adds a dummy for whether a firm is a conglomerate. Row 10 adds the log of firm sales. Row 11 adds a dummy for S&P 500 index membership. Statistical significance at the 5% and 1% levels indicated by ** and *** respectively.

	RATIO	Future region income growth	Future firm ROE	Future firm sales growth	Misc.
1. Baseline specification	−.136*** (.034)				
2. Add region per capita income	−.127*** (.024)				−.138 (.389)
3. Add region population density	−.003 (.061)				−.283*** (.102)
4. Add future region income growth	−.140*** (.038)	.031 (.152)			
5. Add future firm ROE	−.137*** (.038)		.040** (.012)		
6. Add future firm sales growth	−.128*** (.040)			.124*** (.010)	
7. Add all future controls	−.129** (.041)	−.010 (.135)	.037** (.012)	.106*** (.010)	
8. Remove dominant industries	−.114** (.042)				
9. Add conglomerate dummy	−.159** (.061)				−.073** (.022)
10. Add log sales	−.139*** (.032)				.030*** (.006)
11. Add S&P 500 indicator	−.128*** (.038)				.268*** (.044)

(i.e., book value per capita) and that there is not enough variation in the denominator (per-capita income) to isolate its separate contribution. Such a conclusion is not surprising, given the variance decomposition of RATIO discussed above.

In Row 3, we add to the baseline specification region population density. This variable not only attracts a significant negative coefficient, it also completely drives out RATIO. How should one interpret this result? We can imagine two possible views. On the one hand, it can be argued that there is no a priori theoretical reason for population density to go in these regressions. Thus to the extent that it enters significantly, it must be because it effectively cleans up a measurement error problem in our RATIO variable. Recall that in the numerator of RATIO, we have the book value of firms *headquartered* in a given region. However, it is entirely possible that what matters for local bias is not merely the location of a firm's headquarters, but rather the extent of its operating presence (e.g., major manufacturing plants, large R&D campuses) in a given regional economy. Moreover, it may also be that a region's population density better captures the aggregate presence of publicly listed firms than does the book value of firms that are nominally headquartered there; the above-discussed case of Delaware provides a stark illustration of this kind of measurement error.

Under this measurement-error interpretation, it is neither surprising, nor bad news for our theory, that population density takes out RATIO in a simple horse race. A more problematic alternative is that population density is a proxy for some other omitted factor that does legitimately belong in the regression. For example, it may be that regions with low population density have the greatest potential for future growth, which would naturally translate into higher expected cashflows for the firms located there. Fortunately, it is possible to address this sort of alternative hypothesis directly, which we do in Rows 4–7. In Row 4, we add to the baseline specification a term for future region-level income growth, defined as the rate of growth of total region income over years t+1 through t+3.[8] This variable attracts a positive but insignificant coefficient, and has no impact on the RATIO variable. In Rows 5 and 6, we add in turn to the baseline regression average future firm ROE, and future firm sales growth, again measured in each case over years t+1 through t+3. While both of these future firm-level variables have strong positive effects on stock prices, neither appreciably alters the estimated coefficient on RATIO.

Finally, in Row 7, we add all three future-growth terms to the regression simultaneously. Even in this case, the coefficient on RATIO is barely changed, at −0.129. Overall, these variations lead us to conclude that population density is probably not proxying for any kind of directly value-relevant factor—at least not one that shows up in future cashflows.

In Row 8, we return to our baseline specification, but drop all observations corresponding to those firms which belong to "dominant" industries in their region. More precisely, we drop any firm whose two-digit SIC industry accounts for more than 10% of the total book value in a region. (We continue to keep these firms for the purposes of calculating the RATIO variable, however.) The idea here is that if there is still some relevant uncontrolled-for factor at the regional level, it is likely to have more of an effect on dominant-industry firms. For example, suppose that—in the spirit of a multifactor APT—there is an extra risk premium on stocks that are heavily exposed to auto-industry risk. It just so happens that Michigan has an above-average value of RATIO over the full sample period. So one might conceivably argue that perhaps Michigan firms have relatively low market-to-book values not because of a RATIO effect, but because of their high loading on auto-industry risk. By eliminating dominant-industry firms, we throw out any auto firms that happen to be located in Michigan, which should tend to mitigate this type of problem.[9] As it turns out, this adjustment has little effect on our results; the coefficient on RATIO falls only slightly, to −0.114.

In Rows 9–11, we experiment with three other controls that might be expected to have some impact on market-to-book ratios: a conglomerate dummy, the log of firm sales, and an S&P 500 index dummy. None of these controls makes any appreciable difference to the coefficient on the RATIO variable.[10]

[8] We have experimented with variations on the timing—e.g., going 5 or 10 years out when measuring future income growth—with little difference to the results.

[9] When applied to Michigan, our dominant-industry screen excludes the two-digit SIC industry described as "transportation manufacturing," which includes automakers.

[10] The coefficient on the S&P dummy is surprisingly large, at 0.268. But note that this does not imply that the causal effect of S&P inclusion on stock prices is on the order of 27%. It may just be that a high market capitalization is a criterion for S&P inclusion, above and beyond a high book value.

4.5. Alternative measures of RATIO

In Table 7, we try measuring the RATIO variable in different ways. First, in Panel A, we keep all else the same as in the pooled specification in Table 5, but we calculate RATIO at the state level instead of at the Census-region level. This leads to a coefficient of −0.049, with a t-statistic of 2.04. The point estimate implies that if we move from the state with the fourth-highest sample-average value of RATIO (New York, at 0.92) to the state with the fourth-lowest value (Montana, at 0.06) the implied increase in the stock price is 4.3%. Thus the state results are qualitatively similar to those based on Census regions, albeit not quite as strong either statistically or economically. As discussed above, this attenuation of the results is likely due to a combination of two factors. First, as shown in Table 1, investors' preferences for local stocks extend beyond the confines of their home states, and indeed, in a population-weighted sense, the large majority of excess demand associated with local bias comes from outside the state in which a given firm is headquartered. Second, the state-level values of RATIO can be quite noisy, with some of this noise reflecting what is for our purposes measurement error.

Consistent with this reasoning, if we run the regressions at the even finer MSA level, the results (not shown) are completely insignificant. We view this outcome as unsurprising, given the arguments above. On the measurement-error front, the RATIO variable becomes extremely volatile when we compute it at the MSA level. For example, although the sample average value is 0.62, we observe cases like Fayetteville-Bentonville, Arkansas, the MSA that is home to Wal-Mart, where the 2005 value of RATIO is 5.17.

In Panel B of Table 7, we redo both the Census-region and state regressions with cashflow scaling, rather than book-value scaling. This entails two modifications. First, RATIO is now a region's total cashflow divided by its total net income. And second, the dependent variable is now the log of a firm's price-to-cashflow ratio, rather than the log of its market-to-book.[11] As it turns out, the results are very similar to those with book-value scaling, both in terms of statistical significance and economic magnitudes. For example, at the Census region level, the coefficient of −0.555 (t-statistic of 3.99) implies that as we move from the Middle Atlantic to the Deep South, holding all else equal, a firm's stock price goes up by 6.7%; this can be compared to the value of 7.9% from the specification with book-value scaling.

The fact that our results are robust to cashflow scaling is especially helpful in addressing the following issue. As described above, one might worry that the use of book values is problematic to the extent that certain physical assets (e.g., land, or buildings) are more expensive in, say, the Middle Atlantic than the Deep South. If so, this could

[11] Two details about these regressions are worth noting. First, any firms with negative values of cashflow are excluded. Second, given that the left-hand-side variable now includes a measure of profitability, we drop ROE from the right-hand-side of the regressions.

Table 7
RATIO and stock prices, alternative measures of RATIO

The entries are the coefficients on the RATIO variable using the pooled regression specification from Table 5. The first column of both panels shows the estimates when RATIO is calculated at the Census-region level. The second column of both panels uses RATIO calculated at the state level. The two panels vary both the definition of RATIO, and the dependent variable in the regressions. Panel A uses the baseline specification, in which RATIO is a region's total book value divided by its total net income, and in which the dependent variable is the log of the firm's market-to-book ratio. Panel B replaces book value with cashflow, so that RATIO is the region's total cashflow divided by its total net income, and the dependent variable is the log of the firm's price-to-cashflow ratio. Statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and *** respectively.

	1. Census regions	2. States
Panel A: Book-value scaling		
RATIO	−.136*** (.034)	−.049** (.024)
Panel B: Cashflow scaling		
RATIO	−.555*** (.139)	−.176* (.094)

Table 8
RATIO and stock prices, electric utilities

The entries are the coefficients on the RATIO variable using the pooled regression specification from Table 5, with a sample that is restricted to firms that are electric utilities. RATIO is calculated at the Census-region level. The two columns vary both the definition of RATIO, and the dependent variable in the regressions. The first column uses book-value scaling, so that RATIO is a region's total book value divided by its total net income, and the dependent variable is the log of the firm's market-to-book ratio. The second column uses cashflow scaling, so that RATIO is the region's total cashflow divided by its total net income, and the dependent variable is the log of the firm's price-to-cashflow ratio. Statistical significance at the 10% level indicated by *.

	1. Book-value scaling	2. Cashflow scaling
RATIO	−.152* (.080)	−.495* (.263)

potentially induce a negative relationship between a firm's market-to-book, and a RATIO variable whose numerator is also based on book values. However, with cashflow scaling, neither our dependent variable nor RATIO is influenced by the book value of physical assets, which alleviates this concern.

4.6. Evidence from electric utilities

One potential objection to our results thus far goes as follows. Firms' locations are the product of an endogenous choice, and even within an SIC-code-defined industry, there may be firms with different technologies that imply different optimal locations. For example, a firm whose strategy relies heavily on the human capital of computer scientists is presumably more likely to locate where such scientists are abundant, say in the Boston area or Silicon Valley. In contrast, a firm whose technology is relatively land-intensive is more likely to go where real estate is cheap. To the extent that these factors are also correlated with valuations—e.g., human-capital-intensive firms trade at a discount—our inferences could be affected.

In an effort to confront this problem, we re-run our baseline regressions on a subsample of electric utility firms. Our premise here is twofold. First, electric utilities are likely to have relatively homogeneous production technologies across different parts of the country. Second, the fact that there are effectively prohibitive transport costs in this industry implies that the endogenous location-selection effect is unlikely to be at work. Simply put, each region of the country has to have its own locally-based utilities; there is no scope for all firms in the industry to move, e.g., to the Deep South because land or unskilled labor there is cheaper.

Our sample of electric utilities is drawn from firms in SIC codes 4911 ("establishments engaged in the generation, transmission and/or distribution of electric energy for sale") and 4931 ("establishments primarily engaged in providing electric services in combination with other services, with electric services as the major part though less than 95% of the total"). There are roughly 100 such firms in the sample in any given year, with some year-to-year variation.[12] And, as expected, these firms are widely distributed across different regions of the country.

Table 8 presents the results that emerge when our baseline pooled regression specification is applied to the electric-utility subsample. Column 1 shows that when using book-value scaling, the coefficient on RATIO at the Census-region level is −0.152, which is a bit larger in absolute magnitude than the corresponding estimate of −0.136 for all firms. This coefficient implies a stock-price impact of 8.9% when moving from the Middle Atlantic to the Deep South. Column 2 shows that when using cashflow scaling, the coefficient on RATIO at the Census-region level is −0.495, which is a little smaller than the corresponding figure of −0.555 for all firms. While the estimates are in both cases less precise than for the full sample of firms—the *t*-statistics in the two regressions are 1.90 and 1.88 respectively—the similarity of the coefficients to their full-sample values would appear to dispel the claim that our previous findings are the product of some sort of endogenous location-selection mechanism.[13]

4.7. Firms that switch regions

In principle, there is another way to address the concern that systematically different types of firms tend to locate in different parts of the country: one can try to identify the effect of RATIO based on within-firm variation, i.e., by looking at firms that move their headquarters

[12] There are 94 electric utilities in the sample in 1970, 109 in 1975, 110 in 1980, 113 in 1985, 105 in 1990, 98 in 1995, 74 in 2000, and 67 in 2005.

[13] If we re-run things at the state level, the coefficient on RATIO for electric utilities is again negative with both book-value and cashflow scaling. However, neither estimate is statistically significant. This may be due to the simple fact that with such a small number of firms in the utility sample, there are many states (roughly ten in a typical year) with no observations, which further compromises the already-reduced power of the state-level tests.

from one Census region to another at some point in the sample period. Unfortunately, in spite of its conceptual appeal, the power of this approach is limited by a small number of observations. We have found just 23 such "switchers" in the Compustat database with sufficiently complete accounting information that we can include them in our regressions.

Nevertheless, in spite of the small-sample issue, we have run the following regression. For a panel consisting of the 23 switchers, we regress the log of market-to-book against ROE, R&D-to-sales, a dummy for reporting R&D expenditures, and three new variables: AFTER, CHANGE, and the interaction term AFTER∗CHANGE. AFTER is a dummy that takes on the value one in all periods after a firm has changed Census regions. CHANGE is the change in the firm's RATIO associated with its move, that is, the difference between RATIO in its new region and its old region, measured in the year of the move. The key coefficient of interest is that on the interaction term, AFTER∗CHANGE, which tells us how much the move-driven change in RATIO alters the log of market-to-book. Again, note that in this specification, we are now identifying only off of the within-firm time-series changes in RATIO caused by firms' moves.

The point estimate for the AFTER∗CHANGE coefficient in this regression is −0.197. This is in the same ballpark as the estimate of −0.136 from our baseline pooled specification, which is reassuring. Not surprisingly, however, the standard error, at 0.410, is too large for us to draw any statistically firm conclusions.

4.8. The interaction of RATIO and visibility

We now turn to our tests of Hypothesis 2, which suggests that the RATIO variable should have a stronger effect on the prices of less visible firms. To operationalize this hypothesis, we begin by constructing two low-visibility proxies. The first is a dummy that takes on the value one if a firm's size (as measured by its sales) falls outside the top quartile in any given period. The second is a dummy that takes on the value one if a firm's residual number of shareholders falls outside the top quartile in any given period, with the residual based on a regression of the log of the number of shareholders against the log of firm sales. This second low-visibility proxy is, by construction, orthogonal to the first. Moreover, it is more closely linked to our theory, in which visibility is literally defined in terms of whether a firm has a broad (national) or narrow (local) base of shareholders.

Next, in Table 9 we take our baseline pooled specification at the Census-region level, and add two new terms: the low-visibility dummy, and the interaction of this dummy with RATIO. We do this both with the size-based measure of visibility (Row 1 of table), and the measure based on the residual number of shareholders (Row 2 of the table). In both cases, our interest is in the interaction term, which we predict will attract a negative coefficient.

As can be seen, the interaction terms are significantly negative for both low-visibility proxies. Moreover, the associated economic magnitudes are substantial. The coefficients in Row 1 imply that for firms in the top quartile of the size distribution, moving from the Middle Atlantic to the Deep South leads to a 4.1% increase in stock prices, while for firms outside of the top quartile the effect is more than doubled, at 9.9%. The coefficients in Row 2 imply that for firms in the top quartile of the residual-number-of-shareholders distribution, moving from the Middle Atlantic to the Deep South leads to a 3.2% increase in stock prices, while for the remaining, less-visible firms the effect is three times as large, at 9.7%.

Table 9
Interactions of RATIO with measures of visibility

The regressions in this table begin with the pooled specification reported in Table 5, and add two independent variables: (i) a low-visibility dummy (not shown in the table); and (ii) the interaction of the low-visibility dummy and RATIO. RATIO is measured at the Census-region level. In the first row, the low-visibility dummy takes on the value one if the firm's sales rank below the top quartile. In the second row, the low-visibility dummy takes on the value one if the firm's residual number of shareholders ranks below the top quartile. The residual number of shareholders is based on a regression of the log of the number of shareholders against the log of firm sales. Statistical significance at the 5% level indicated by **.

	RATIO	(Low-visibility)* RATIO
1. Low visibility = sales below top quartile	−.071 (.050)	−.097** (.048)
2. Low visibility = residual number of shareholders below top quartile	−.056 (.051)	−.109** (.052)

5. Discussion

5.1. Implications for expected returns and arbitrage

We have framed our entire empirical analysis in terms of the *level* of stock prices. But as a logical matter, our theory makes a corresponding set of predictions about stock returns. In particular, a stock located in a region with a low value of RATIO should have a high price precisely because it has a low expected return. Moreover, casting the tests in terms of expected returns would seem to have the added advantage of more fully controlling for unobserved heterogeneity in future cashflows. For example, if stocks located in the Deep South have higher prices, there is always the worry that this is in part because there is some missing Deep-South factor (e.g., hidden long-run growth potential) that will ultimately lead to higher cashflows and hence justify the higher prices. If, however, Deep-South stocks have persistently lower returns, such an alternative hypothesis can be dismissed.

So why not look at returns instead of price levels? The answer is that, since all risk premia in our model are permanent in nature, the sorts of price effects that we have documented translate into very small expected-return differentials—far too small to show up as statistically significant, given the power of such tests. To see this concretely, think of a stock with a price-earnings (P/E) ratio of 20. In a simple perpetuity formula, this corresponds to a value of $(k-g)$ of 5%, where k is the discount rate and g is the growth rate. Now, in line with our empirical estimates, raise the price of the stock by 8%, so that the P/E goes to 21.6. With this realistic calibration, the implied value of $(k-g)$ only falls to 4.63%. In other words, the expected return only drops by 37 basis points per year. This is simply too small an effect to pick up with a standard return-forecasting exercise.

To see this point explicitly, consider the following regression. We run firm stock returns in year t+1 against RATIO measured in year t, as well as a set of size dummies (nothing changes if we also add controls for, e.g., book-to-market or momentum to the regression). Our most precise estimates come if we restrict the sample to electric utility firms, thereby eliminating the noise that comes from industry-level return shocks. Yet even in this best-case scenario, we get a coefficient of 0.0033 on RATIO, with a standard error of 0.0116. The point estimate is of the right sign, and actually quite sensible in magnitude—it implies that as we move from the Deep South to the Middle Atlantic, expected returns go up by 18 basis points per year, or about half of what we would predict based on the simple calibration above. Unfortunately, the point estimate is only about one-quarter the size of the standard error, so it is impossible to make meaningful inferences. And again, this power problem is even worse when we look at the full sample of firms, where the standard error of the RATIO coefficient is roughly doubled.

This logic also sheds light on why arbitrage is unlikely to eliminate the price-level effects that we document. To exploit the pricing discrepancies across regions, an arbitrageur would have to buy the stocks of Middle-Atlantic firms, and short the stocks of Deep-South firms. In doing so, he would incur substantial regional risk—the economy of the Deep South could boom unexpectedly relative to that of the Middle Atlantic, thereby devastating his position—all for an annualized alpha (before transactions costs) on the order of 37 basis points. This hardly seems like an attractive trading strategy. Thus in contrast to other, faster-converging phenomena like medium-term momentum (Jegadeesh and Titman, 1993) or post-earnings-announcement drift (Bernard and Thomas, 1989, 1990), here we have a case where there are economically meaningful price-level effects, but little that would be of interest to a money manager.

This observation about speed of convergence highlights the key difference between our work and that of Coval and Moskowitz (2001), Ivkovic and Weisbenner (2005), and Ivkovic, Sialm, and Weisbenner (2008). These papers show that investors make significant excess returns when trading in local stocks, even after controlling for factors such as size, book-to-market, and momentum. These findings suggest that local investors have short-lived private information about future returns, information that allows them to time their buys and sells better than non-local investors. By contrast, our hypothesis has to do with permanent, unconditional differences in price levels. Thus it should not be surprising that we have less to say than these other papers about short-horizon expected returns. It is presumably for this same reason that the literature on index inclusion effects (initiated by Harris and Gurel, 1986; Shleifer, 1986) always looks at prices, rather than expected returns. To the extent that any inclusion effect is both permanent and modest in magnitude, it makes little sense to test the hypothesis that, say, S&P 500 stocks have lower expected returns than non-S&P 500 stocks.

5.2. International asset pricing with segmented markets

There is a clear parallel between our work and the literature on international asset pricing in the presence of segmented markets. One major branch of this literature adopts a CAPM perspective, and asks whether expected returns on stocks in a given small country are driven by their betas with respect to the home-country market portfolio, or their betas with respect to the world market portfolio.[14] Typically, the home-country and world equity premia are taken as exogenous in these papers, and little effort is devoted to understanding their determinants. In contrast, we completely ignore beta considerations: in our model, both the local-market and national-market betas of all stocks are effectively set equal to one. Thus our analysis can be thought of as focusing exclusively on the determinants of average local-market equity premia.[15]

It is natural to wonder what implications, if any, our results have for cross-country differences in asset prices. At a general level, they would certainly seem to suggest that local-market supply and demand factors can have meaningful consequences for price levels.[16] At the same time, it would probably be naïve to run cross-country versions of our regressions—with country-wide analogs to the RATIO variable—and expect to get similar results.

One obvious complicating factor has to do with differences across countries in financial development. For example, a country with weak investor protection is likely to have both lower stock prices (LaPorta, Lopez-de-Silanes, Shleifer, and Vishny, 2002) as well as fewer publicly listed firms (LaPorta, Lopez-de-Silanes, Shleifer, and Vishny, 1997), and hence a lower value of RATIO. To the extent that it is not possible to control perfectly for the degree of investor protection, this effect will tend to obscure the negative relationship between RATIO and stock prices that we observe in the U.S. data.

Of course, there are some noteworthy institutional differences—in, e.g., zoning and bankruptcy laws—even across different U.S. regions. However, we have seen that within the U.S., cross-region variation in RATIO is driven largely by variation in population density. And we have

[14] Notable papers include Stulz (1981), Errunza and Losq (1985), Eun and Janakiramanan (1986), Jorion and Schwartz (1986), Wheatley (1988), Hietala (1989), Bailey and Jagtiani (1994), and Chari and Henry (2004).

[15] In this regard, we are posing a question somewhat analogous to Bekaert and Harvey (2000), and Henry (2000). Both of these papers show that on average, prices go up—and equity premia presumably go down—when an emerging stock market is opened up to foreign investment.

[16] Further support for this hypothesis comes from recent work by Braun and Larrain (2006). They show that large initial public offerings (IPO) in emerging markets tend to depress the prices of firms in those industries that covary highly with the industry of the firm conducting the IPO. This is consistent with the same sort of segmented-markets supply effect that we focus on.

argued that this latter sort of variation is plausibly exogenous with respect to the level of stock prices. In contrast, when one looks across countries, institutional factors are likely to play a bigger role in influencing RATIO, and these other factors may well not be exogenous with respect to stock prices.

6. Conclusions

The basic message of this paper is a simple one: like many other goods and services, stocks have prices that can be materially influenced by local supply and demand conditions. Just as one would expect the price of a hotel room to be lower in a city where hotel rooms are plentiful, so too is the price of a firm's stock lower if it is located in a region where it must compete for investors' dollars with many other nearby firms. The magnitude of this effect is surprisingly large, especially among smaller, less-visible firms, where the implied price differentials across Census regions are as high as 10%.

In closing, we should stress two implications that our analysis *does not* have. First, there is nothing in our results that suggests that any given firm can be made better off—in the sense of generating a higher stock price—by moving to a region with a lower value of RATIO. Recall that every one of our specifications looks at the effect of the RATIO variable *holding fixed firm profitability*, as measured by ROE. And it is obviously unlikely that a typical firm located in a high-RATIO state like New York could move to a low-RATIO state like West Virginia without adversely affecting its profitability.

Second, in spite of the relatively large stock-price effects that we document, there is little here of interest to would-be arbitrageurs. Given that location exerts a permanent influence on expected returns, it takes only a small rate-of-return wedge to generate the sorts of price-level differentials that we see in the data. Thus, any arbitrage strategy based on our findings is likely to have a very small alpha relative to the associated risks and transactions costs. Indeed, it is for precisely this reason that even our most aggressive price-level estimates can be defended as economically plausible, since they do not suggest any easily exploitable arbitrage opportunities.

References

Bailey, W., Jagtiani, J., 1994. Foreign ownership restrictions and stock prices in the Thai capital market. Journal of Financial Economics 36, 57–87.

Barber, B., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance 55, 773–806.

Bebchuk, L., Cohen, A., 2003. Firms' decisions where to incorporate. Journal of Law and Economics 46, 383–425.

Bekaert, G., Harvey, C., 2000. Foreign speculators and emerging equity markets. Journal of Finance 55, 565–613.

Bernard, V.L., Thomas, J., 1989. Post-earnings announcement drift: delayed price response or risk premium? Journal of Accounting Research 27, 1–48.

Bernard, V.L., Thomas, J., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. Journal of Accounting and Economics 13, 305–340.

Braun, M., Larrain, B., 2006. Do IPOs affect the prices of other stocks? Evidence from emerging markets. Unpublished working paper, Pontificia Universidad Catolica de Chile.

Chan, L., Lakonishok, J., Sougianis, T., 1999. The stock market valuation of research and development expenditures. Unpublished working paper, National Bureau of Economic Research.

Chari, A., Henry, P., 2004. Risk sharing and asset prices: evidence from a natural experiment. Journal of Finance 59, 1295–1324.

Cooper, I., Kaplanis, E., 1994. Home bias in equity portfolios, inflation hedging and international capital market equilibrium. Review of Financial Studies 7, 45–60.

Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: local equity preference in domestic portfolios. Journal of Finance 54, 2045–2073.

Coval, J.D., Moskowitz, T.J., 2001. The geography of investment: informed trading and asset prices. Journal of Political Economy 109, 811–841.

Errunza, V., Losq, E., 1985. International asset pricing under mild segmentation: theory and test. Journal of Finance 40, 105–124.

Eun, C.S., Janakiramanan, S., 1986. A model of international asset pricing with a constraint on the foreign equity ownership. Journal of Finance 41, 897–914.

Fama, E.F., MacBeth, J.D., 1973. Risk, return and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.

French, K., Poterba, J., 1991. Investor diversification and international equity markets. American Economic Review 81, 222–226.

Grinblatt, M., Keloharju, M., 2000. The investment behavior and performance of various investor-types: a study of Finland's unique data set. Journal of Financial Economics 55, 43–67.

Harris, L., Gurel, E., 1986. Price and volume effects associated with changes in the S&P 500 list: new evidence for the existence of price pressures. Journal of Finance 41, 815–829.

Henry, P., 2000. Stock market liberalization, economic reform and emerging market equity prices. Journal of Finance 55, 529–564.

Hietala, P., 1989. Asset pricing in partially segmented markets: evidence from the Finnish market. Journal of Finance 44, 697–718.

Huberman, G., 2001. Familiarity breeds investment. Review of Financial Studies 14, 659–680.

Ivkovic, Z., Weisbenner, S.J., 2005. Local does as local is: information content of the geography of individual investors' common stock investments. Journal of Finance 60, 267–306.

Ivkovic, Z., Sialm, C., Weisbenner, S.J., 2008. Portfolio concentration and the performance of individual investors. Journal of Financial and Quantitative Analysis, forthcoming.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Jorion, P., Schwartz, E., 1986. Integration vs. segmentation in the Canadian stock market. Journal of Finance 41, 603–614.

LaPorta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 1997. Legal determinants of external finance. Journal of Finance 52, 1131–1150.

LaPorta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 2002. Investor protection and corporate valuation. Journal of Finance 57, 1147–1170.

Moskowitz, T., Vissing-Jorgensen, A., 2002. The returns to entrepreneurial investment: a private equity premium puzzle? American Economic Review 92, 745–778.

Petersen, M.A., 2005. Estimating standard errors in finance panel data sets: comparing approaches. Unpublished working paper, Northwestern University.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Shleifer, A., 1986. Do demand curves for stocks slope down? Journal of Finance 41, 579–590.

Stulz, R., 1981. On the effects of barriers to international investment. Journal of Finance 36, 923–934.

Van Nieuwerburgh, S., Veldkamp, L., 2007. Information immobility and the home bias puzzle. Unpublished working paper, New York University.

Wheatley, S., 1988. Some tests of barriers to international investment. Journal of Financial Economics 21, 177–212.

Zhu, N., 2003. The local bias of individual investors. Unpublished working paper, University of California, Davis.