FAME

FINANCE & ACCOUNTING MEMOS

jagazine

e E-Book   www.fame-jagazine.com   Free Web Online

Issue 1, 2014

Are Stocks Really Less Volatile in the Long Run?

Why Do Investors Trade?

Family-Controlled Firms and Informed Trading

Hidden Information in Insider Trading

Short-Selling Bans Around the World

Systemic Risk and the Refinancing Ratchet Effect

and more

Launch Sponsor: Research Affiliates

About the Cover

Quentin Matsys (1466–1529): The Moneylender and his Wife. Flanders, 1514. This painting in the Louvre represents a period of flourishing merchant trade in Flanders. The painting is well-known amongst economists and accountants alike. The money changer focuses on carrying out his trade, while his equally-engrossed wife looks at the coins holding a religious book in her hands. Does this mean she is praying?

Advice to the (e-)reader

FAM𝑒 is available in pdf form, html form, and epub form. Each format has advantages and disadvantages:

- Pdf is the best format for printing. The pdf layout is fixed. FAM𝑒 can control both fonts and the placement of material. Pdf is not bad online, either, with very high-quality font rasterization on all platforms. However, it cannot re-flow the text according to the reader's font-size preference and viewing device. If you want to view a pdf page that is twice as wide, you need a canvas that is twice as wide.

- Html is the best format for on-line viewing on a high-speed-internet-connected *computer*. It is usually easy to change the browser font-size (using ctrl-minus or ctrl-plus) with a concomitant re-flow of the document. Unfortunately, the most common *tablet* web browsers (like Safari on iOS, Chrome on Android, and Internet Explorer on Windows-8) make font-size changes difficult. Pinching html text in tablets usually zooms into the document (just as it does in a pdf file), instead of re-flowing the text.

- Epub is the best format for tablets. Like pdf and unlike html, it works well for viewing on non-Internet connected devices, too. (Epub is basically pre-packaged html without javascript.) Unfortunately, many e-book readers are still "immature" (buggy). FAM𝑒 has been tested to work well on Apple iBooks, Firefox's epubreader, and Adobe's Digital Editions. calibre and the Android Universal Book Reader mostly work, but have several bugs. (Amazon makes it intentionally difficult to transfer content that was not obtained through the Amazon store onto the Kindle reader.) Note that external links cannot be made to work reliably inside an ebook reader—epub is a self-contained package format. Some e-book readers will open a web browser; others will not.

- Source contains the LaTeX-dialect source from which all of the above have been created. This is of interest only to connoisseurs of typesetting systems.

In both browser html and ereader epub format, it is the client program and not the FAM𝑒 team that handles font-size, line-breaking, and page-breaking. This is often but not always good. For example, tables and figures may be broken at *very* inopportune spots. Fortunately, when the reader resizes the content on a strange-looking page, the page often suddenly looks great. (linux users: please install msttcorefonts or ttf-mscorefonts-installer.)

Academic articles often include tables that require a wide page for comprehension. For this reason, FAM𝑒 articles are not well-suited to reading on small-screen devices. A 10-inch diagonal high-resolution tablet screen is recommended, although a 7-8-inch high-resolution diagonal screen size may be acceptable. (Also, note that wide-screen is great if your e-reader does not decide to abuse it for two-column reformatting, in which case it becomes narrow-screen.) Please, do not think about reading FAM𝑒 on a 3.3-inch diagonal iphone. It would be an exercise in frustration.

Small text font-sizes often work better with wide tables, but large font-sizes often look better on the main text in the absence of tables and figures. If you can easily switch font-size on the fly, you can get the best of both worlds.

www.fame-jagazine.com

The first issues of FAMe are still distributed in print. This will not last forever. We are planning to move to a complete on-line model sometime in the future.

Submissions, advertising, and support

Advertisements and support

For advertisement inquiries, please email fame.jagazine@gmail.com.

FAMe's goal is to broaden the impact of academic finance and accounting research. Creating an issue of FAMe costs about $50,000, largely financed out of pocket by ourselves. We depend on the goodwill of authors, readers, and supporters. We cannot spread costs over a large subscriber base—we are not *The Economist*.

To help underwrite our efforts, please consider purchasing a $50 or $100 annual subscription, especially if you can cover this subscription expense from your research budget. There is a paypal button on our website (click here). The paypal receipt should make it easy to be reimbursed as a research expense. The **library** subscription fee is the same price as the individual subscription fee (and paid the same way). The $100 subscription fee entitles libraries to make the content available to their patrons online.

Instructions to authors

To submit a FAMe version of your article recently published or forthcoming in a "pre-approved" finance journal, please send your proposed short memo to: famesubmission@gmail.com. For papers from accounting journals, please use famesubmission+a@gmail.com, instead. For more information and detailed instructions on how to prepare a MeMO, please refer to FAMe guidelines for authors on the www.fame-jagazine.com website for more information. Read our most recent issue to understand our "flavor"—we are still evolving, too. Detailed instructions for accepted submissions are posted at http://www.fame-jagazine.info/.

Contact

If you have contacted us at one of the above email addresses (fame.jagazine@gmail.com), e.g., famesubmission@gmail.com, and have not heard back from us, please feel free to prod us again at ivo.welch@gmail.com and bhagwan@anderson.ucla.edu.

Featured MeMos

Bhagwan Chowdhry, Executive Editor

Editorial: Who wouldn't like FAMe?

The top academic finance and accounting journals collectively publish about 500 research articles every year. Even subscribers do not read most of them. Exposure of scholarly research to practitioners and policy makers is even more limited and sporadic. Once in a while, a journalist from The Economist, the Wall Street Journal, or the Financial Times features a research paper, generating a flurry of short-lived interest from many, a huge number of downloads, and fifteen minutes of fame for the authors. Most papers never receive such attention.

This is about to change.

Welcome to the inaugural issue of Finance & Accounting Memos (FAMe). FAMe is inviting authors of articles accepted for publication in top finance and accounting journals (and occasionally beyond) to write short versions of their articles in language that makes the main ideas and results more accessible to broader audiences. FAMe synopses should not be cut-and-paste from the abstract, introduction and conclusion of the original article but distill the paper's key propositions. This usually entails a few equations, numerical examples, key tables, or figures.

FAMe is a "cross-over" journal/magazine = **jagazine**, clearly academic in nature, yet understandable by interested and smart non-academics. We strongly request of our readers that FAMe *not* be cited. All citations should go to the original journal articles instead in order to emphasize where the real research was published.

We believe that FAMe will make our academic journals, societies, and profession better. Papers, authors, and journals will garner more visibility, readership, and citations. Readers will find it easier to keep up with more research from beyond their own specific fields of interest. Doctoral students will find it easier to digest more current research quickly and efficiently. MBA students and practitioners will find academic research more accessible and relevant. (We hope professors will assign FAMe versions in their classes.) Journalists and policy makers will find it easier to access more research to guide public debate.

Of course, FAMe will only succeed if all of us collectively see value in it, contribute to it, and provide suggestions for improvement. Many thanks to all the authors who contributed to the inaugural issue—please spread the word and encourage your colleagues to write and submit for future issues of FAMe.

Many people helped make FAMe possible. We thank Ken Singleton (JF), Campbell Harvey (JF), Sheridan Titman (AFA), David Hirshleifer (RFS), Michael Weisbach (RFS), Wayne Ferson (RAPS), Paolo Fulghieri (RCFS), Matthew Spiegel (SFS), Bill Schwert (JFE), Paul Malatesta (JFQA), Hank Bessembinder (JFQA), Jarrad Harford (JFQA), Stephen Brown (JFQA), Ray Ball (JAR), Jerry Zimmerman (JAE), Ross Watts (JAE), Richard Sloan (RAST), Harry Evans (TAR) for their encouragement, support, and believing in what FAMe is trying to accomplish. Zac Rolnik and Mike Casey (NOW Publishers) helped brainstorm many useful suggestions for publishing FAMe. Ani Adzhemyan worked on the first version. Swati Desai and Lily Qiu were indispensable in FAMe production, art design and helped curate the paintings.

David Aboody, Amit Goyal and Ivo Welch as co-editors worked constantly with me on all aspects of FAMe. FAMe Associate Editors generously contributed their time and suggestions in reading the submissions and helped many authors rewrite their submissions. Without our day jobs and intellectual grounding at the Anderson School, the academic home of many of our editors and myself, we could not have undertaken our new venture. And finally, FAMe would not have seen the light of day, had it not been for Ivo Welch who not only prodded me to realize an idea that I had been discussing with many of you for years, but also underwrote the cost of developing and producing FAMe. In addition, he spent many hours designing and coding so that FAMe could be produced in formats suitable for printing (pdf), reading on tablets and e-readers (epub), and online web formats simultaneously. On behalf of myself and the profession, I offer my gratitude to Ivo.

Enjoy the inaugural issue of FAMe. FAMe is available both in print format and online, and is optimized for large-screen tablets. Incidentally, all paintings are generously hyperlinked—as are the FAMe memos—an invitation to explore deeper. Go look!

Bhagwan Chowdhry

Executive Editor

And do not forget to download issue #2 (corporate finance) at

www.fame-jagazine.com

It's available in pdf, html, and ebook format!

Lubos Pastor and Robert F. Stambaugh

Are stocks really less volatile in the long run?

Watch Lubos Pastor's talk

According to conventional wisdom, stock returns are less volatile over longer investment horizons. The idea is that bull and bear markets partially offset each other, reducing long-horizon variance. This reasoning, supported by historical estimates of volatility, is often invoked to justify generous stock allocations for long-horizon investors.

Our JF article reaches the opposite conclusion: stocks are actually more volatile over longer investment horizons. The key to our conclusion is that we take an investor's perspective. Instead of calculating backward-looking historical estimates of volatility, we calculate forward-looking measures of volatility that are relevant to investors.

Recognizing uncertainty

Investors are uncertain about the extent to which future stock returns will behave similarly to historical estimates. Our measure of volatility incorporates this "parameter uncertainty," whereas historical volatility does not. The forward-looking volatility we calculate, commonly referred to as predictive volatility in Bayesian statistics, is the relevant volatility from an investor's perspective. We find empirically that the U.S. stock market's predictive volatility exceeds its historical volatility, especially at long investment horizons. Moreover, predictive volatility increases with the investment horizon, unlike historical volatility, which exhibits the opposite pattern.

A key parameter governing future stock returns is the equity premium, μ_t, the stock market return one should expect in year $t + 1$ relative to a riskless investment. Even after observing two centuries of stock market returns, investors are uncertain about the current μ_t, as well as how it might change in the future. To compute predictive volatility, we must specify how μ_t can change over time. It is commonly assumed that μ_t depends on the investment environment via a set of observable predictors, x_t. Specifically, $\mu_t = a + b' \cdot x_t$. This is a useful assumption in many applications, but we relax it here because it understates the uncertainty faced by an investor assessing the volatility of future returns. No investor can be certain that μ_t is perfectly captured by x_t. It seems much more likely that a set of observed predictors is imperfect, in that $\mu_t = a + b' \cdot x_t + \pi_t$, where π_t can be

viewed as an unobservable predictor at time t. To admit such predictor imperfection, we employ a predictive system, an econometric model that we developed in Pastor-Stambaugh (JF 2009). The predictive system assumes

$$r_{t+1} = a + b' \cdot x_t + \pi_t + u_{t+1}, \tag{1}$$

where r is the stock market return at time $t+1$, and u_{t+1} is a random error. Recognizing uncertainty due to predictor imperfection is important in reaching our conclusions.

We estimate the predictive system on 206 years of annual real U.S. stock market returns, covering the period 1802 through 2007. We consider three observable predictors (x_t): the aggregate dividend yield on U.S equity, the term spread (i.e., the difference between the long-term high-grade bond yield and the short-term interest rate), and the change in the long-term bond yield. These predictors seem reasonable choices given the various predictors used in previous studies and the information available in the historical data set, for which we are grateful to Jeremy Siegel. All three predictors exhibit significant ability to predict next year's market return.

The main result

The solid line in Figure 1 plots predictive volatility as a function of the investment horizon k. Predictive volatility is the annualized standard deviation of market returns over the following k years, calculated based on all available data at the end of our sample. (Formally, annualized predictive volatility is given by $\text{Std}(r_{T,T+k}|D_T)/\sqrt{k}$, where $r_{T,T+k}$ is the cumulative return between times T and $T+k$, and D_T contains all return and predictor data available at time T.) The figure shows that predictive volatility rises with the investment horizon, from about 17% per year at the one-year horizon to almost 21% per year at the 30-year horizon. Long-horizon stock investors clearly face more volatility than short-horizon investors on a per-year basis.

Figure 1: Stock volatility at different horizons
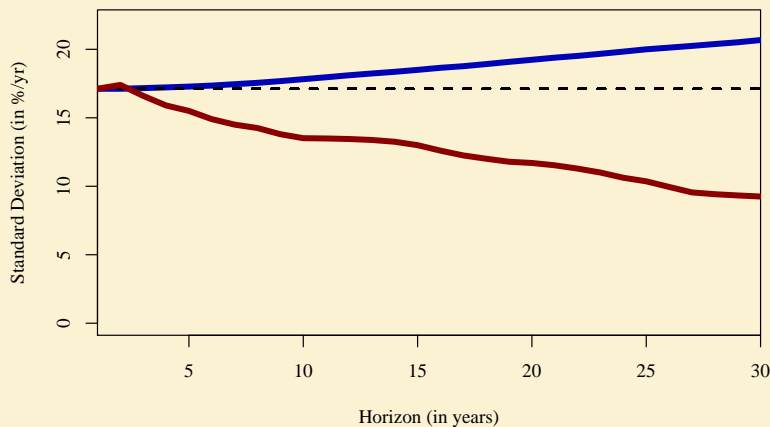
The blue line is the predictive volatility, relevant to investors. The red line is the historical volatility. The horizontal black dashed line is the random-walk benchmark.

For comparison, the dashed line in Figure 1 plots the historical volatility as a function of the investment horizon. Historical volatility at a given horizon k is computed as the annualized standard deviation of cumulative market returns over all k-year periods in our sample. For example, for k = 10 years, we compute the cumulative return for each of the 197 overlapping 10-year periods 1802–1811, 1803–1812, … , 1998–2007, then we calculate the sample standard deviation of the 197 returns, and finally we annualize the standard deviation by dividing it by the square root of 10. The figure shows that historical volatility decreases with the investment horizon, from 17% per year at the one-year horizon to 9.3% per year at the 30-year horizon. The dashed line is the source of the conventional wisdom—historically, stocks have been less volatile in the long run. (The same conclusion obtains based on "conditional" volatility, which conditions on information useful in predicting returns but also ignores parameter uncertainty, e.g., see work by John Campbell and Luis Viceira.)

Where does our main result come from?

To understand the difference between the solid and dashed lines, it is useful to begin by considering a hypothetical world in which stock prices follow a random walk—μ_t is constant over time—and there is no parameter uncertainty. In such a world, annualized predictive volatility does not depend on the investment horizon, as shown by the flat dotted line in Figure 1.

The real world differs from the hypothetical world in two important ways. First, stock prices do not follow a random walk. Instead, they exhibit a certain degree of "mean reversion," in that unexpectedly high returns tend to be followed by a lower μ_t, and low returns tend to be followed by a higher μ_t. Mean reversion pulls long-horizon volatilities down. Second, investors face substantial parameter uncertainty. Investors are uncertain about the current and future values of μ_t as well as other parameters of stock returns such as volatility or persistence. Parameter uncertainty pulls long-horizon volatilities up.

Historical volatility reflects mean reversion but not parameter uncertainty; hence the dashed line is downward-sloping. Predictive volatility reflects both mean reversion and parameter uncertainty. These two forces pull in opposite directions, but our results show that parameter uncertainty prevails; hence the solid line is upward-sloping. That is our main result.

For more intuition, consider the trend from which stock prices randomly depart. Historical volatility is computed around the historical trend, which is known. The future trend is unknown, however, so forward-looking volatility must reflect not only random departures from the trend but also uncertainty about the trend itself. Due to the latter uncertainty, predictive volatility exceeds historical volatility. Moreover, the wedge between the two volatilities increases with the investment horizon. Trend uncertainty does not matter much at short horizons, but it compounds quickly as the horizon lengthens.

We also find that long-run predictive volatility is substantially higher compared to the framework in which predictors are perfect (i.e., when π_t is omitted from equation (1)). Predictor perfection and parameter uncertainty interact—once predictor imperfection is admitted, parameter uncertainty

is more important in general. In particular, when the conditional mean is not observed, learning about its properties is harder compared to the perfect-predictor case.

We show that our main result is robust to various modifications of the basic framework. We consider three econometric models: two predictive systems as well as a predictive regression with an uncertain set of observable predictors. We split the sample in half and run the analysis on both sub-samples. We replace our annual sample with a quarterly sample of post-war returns. We consider different observable predictors. All of these exercises lead to the same basic conclusion: stocks are more volatile over longer horizons from an investor's perspective.

Implications for investors

Our conclusion makes stocks less appealing to long-horizon investors than conventional wisdom would suggest. As a result, buy-and-hold investors who have invested based on historical estimates of volatility might reconsider their stock allocations. We take a closer look at the implications of our findings for investors in target-date retirement funds, which have become very popular recently. These funds follow a pre-determined asset allocation policy that gradually reduces the stock allocation as the target date approaches, with the aim of providing a more conservative asset mix to investors approaching retirement.

We analyze target-date funds using a simple model in which a risk-averse investor can invest in only two assets, the stock market and a real riskless asset. The investor focuses on the final level of wealth achieved at the end of a K-year horizon. (Specifically, the investor maximizes the expected value of utility $W_K^{1-A}/(1-A)$. for end-of-horizon wealth W_K.) The investor commits at the outset to a pre-determined investment strategy in which the stock allocation evolves linearly from the first-period allocation w_1 to the final-period allocation w_K. The investor chooses the values of w_1 and w_K within the $(0,1)$ interval. We allow the investor to save from labor income, which we calibrate to match the observed hump-shaped pattern of labor income over a typical American's life cycle. We obtain the same conclusions in the absence of labor income.

Figure 2 plots the investor's optimal initial and final stock allocations, w_1 (solid line) and w_K (dashed line), respectively, for investment horizons ranging from one to 30 years. We incorporate parameter uncertainty in Panel B but not in Panel A.

Figure 2: Allocations with and without parameter uncertainty
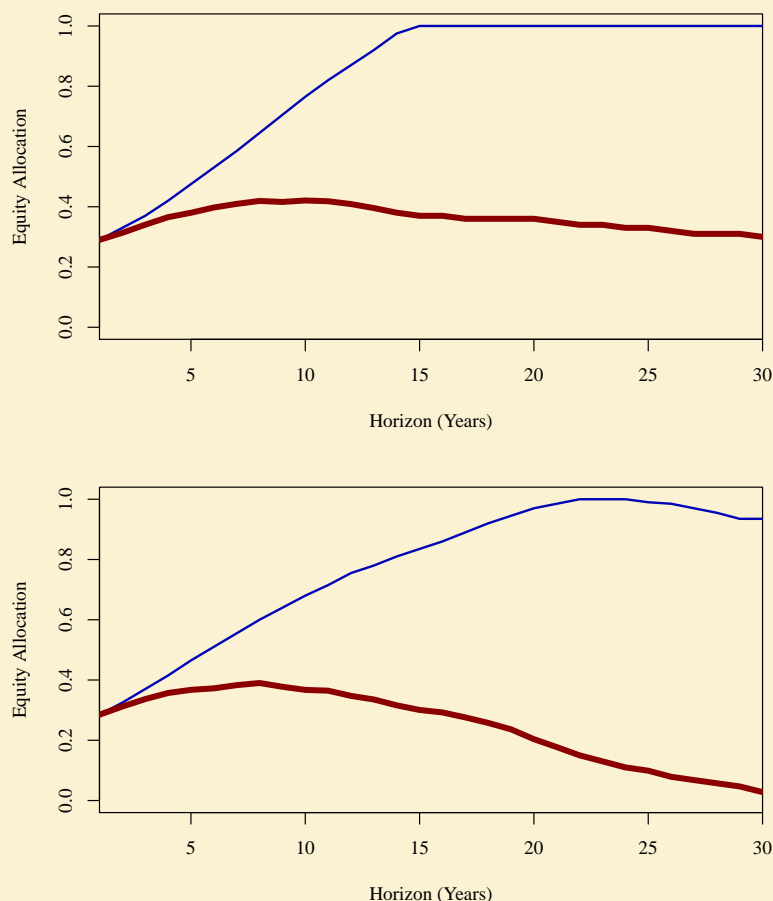
The top graph is the case without parameter uncertainty. The lower graph is the case with parameter uncertainty. The initial allocation is blue. The final allocation is red. For example, for an investor with a 30-year horizon, the optimal equity allocation is about 3% in the absence of parameter uncertainty, much lower than the 30% in its presence.

The optimal allocations in the top graph in Figure 2 are strikingly similar to those selected by real-world target-date funds. The initial allocation w_1 decreases as the investment horizon shortens, declining from 100% at horizons longer than 15 years to about 30% at the one-year horizon, whereas the final allocation w_K is roughly constant at about 30% to 40% across all horizons. Investors in real-world target-date funds similarly commit to a stock allocation schedule, or "glide path," that decreases steadily to a given level at the target date. The final stock allocation in a target-date fund does not depend on when investors enter the fund, but the initial allocation does—it is higher for investors entering longer before the target date. Not only the patterns but also the magnitudes of the optimal allocations in the top panel resemble those of target-date funds. In short, target-date funds seem appealing to investors who ignore parameter uncertainty.

In contrast, the lower graph in Figure 2 shows that target-date funds do not appear desirable if the same investors acknowledge a realistic amount of parameter uncertainty. For short investment horizons, the results look similar to those in Panel A, but for horizons longer than 23 years, both w_1 and w_K decrease with K. For example, an investor with a 23-year horizon chooses to glide from $w_1 = 100\%$ to $w_{23} = 14\%$, whereas an investor with a 30-year horizon glides from $w_1 = 93\%$ to $w_{30} = 3\%$. Parameter uncertainty clearly matters more at longer investment horizons.

Figure 2 shows that parameter uncertainty makes target-date funds undesirable when they would otherwise be virtually optimal for investors who desire a pre-determined asset allocation policy. It would be premature, however, to conclude that parameter uncertainty makes target-date funds undesirable in all settings. Our analysis abstracts from many important considerations faced by investors, such as intermediate consumption, additional risky assets, housing, etc. Our objective is simply to illustrate how parameter uncertainty can reduce the stock allocations of long-horizon investors, consistent with our results about long-horizon volatility.

Titian (1488-1576): Tribute Money. Italy, 1516. The inquiring Pharisees are shocked to hear Jesus telling them to pay taxes to Caesar after reading the inscription on the coin, "Give to Caesar what is Caesar's." Titian painted two versions of the same subject separated by 50 years, one in 1515 and the other one in 1568. Which one do you consider to be the more "mature" version?

Jonathan E. Ingersoll, Jr. and Lawrence J. Jin

Realization utility with reference-dependent preferences

Why do investors buy and sell stocks? The most natural answer is that investors believe that they have new information and want to act on it to maximize their future returns or limit their risk. Alternatively, investors may trade a stock because selling it at a gain gives them a burst of pleasure by confirming that the original decision to buy was correct. Conversely losses may be avoided as they indicate faulty decisions. In academic terms, we call the former type of explanation belief-based and the latter type preference-based. In our paper, "Realization Utility with Reference-Dependent Preferences" (Ingersoll-Jin (RFS 2013)) and this shorter note, we focus entirely on preference-based explanations for trading.

Profits make me happy, losses make me sad...

Of course, there are many preference-based, as well as belief-based, stories that help explain investors' trading behaviors. Investors may trade stocks after a change in family or employment status or due to new objectives in life. They also may trade when their existing portfolios have become unbalanced. Despite such explanations, quite a few empirical facts remain puzzling. For example, academics have no commonly agreed-upon rationale for why individual investors trade excessively even though they underperform passive indices on average. Our general goal is to develop theoretical models to close this gap. Our particular explanation focuses on realization utility as introduced in a dynamic setting by Barberis-Xiong (JFE 2012). Realization utility is based the idea that whatever pleasure or pain is attained from investing comes only when the investor takes the definite action of closing out a position.

...but the feelings diminish

To provide a realization utility framework to explain investors' trading behaviors, an important question is why do investors sometimes sell stock at a loss? If the purpose of trading stock is to enjoy the good feeling of realizing gains, why don't investors always hold on to losses hoping for a later recovery in the stock price? Again one can think of many belief-based and preference-based explanations, but our goal is to provide a unified explanation of both selling at gains and at losses using the idea of realization utility. In order to do so, we incorporate a key ingredient from the

behavioral-economics literature: as the level of overall gains and losses increases, the pleasure and pain brought to investors by an extra dollar of gain and loss tends to decrease. For instance, making an additional one thousand dollar gain on top of a million dollar profit does not bring you as much excitement as does the first one thousand dollars of the profit. This is also true for losing an additional one thousand dollars. Academics commonly call this property diminishing sensitivity, as discussed by Tversky-Kahneman (JRU 1992). Note that this differs from the standard assumption in economic theory that there is increasing sensitivity for larger losses.

In Ingersoll-Jin (RFS 2013), we use a dynamic model to study both realization utility and diminishing sensitivity. Our key finding is that incorporating diminishing sensitivity into a realization utility framework significantly improves the model's match to the empirical facts on trading activities and asset prices. In particular, without diminishing sensitivity, investors never voluntarily sell stocks at a loss. Adding this component, however, generates the model predictions that investors do optimally realize both gains and losses and that the former occurs more frequently.

Our model

In this note, we present a simplified version of the model in our RFS paper to illustrate the key insights. We then discuss how these results help explain some puzzling empirical facts. Finally, we provide thoughts on testing some new predictions.

Consider the following economic setup. An investor purchases a share of a stock. Each period while the share is held, its price increases or decreases by one dollar with probabilities π and $1 - \pi$, respectively. When the investor sells, she gets a burst of realization utility that depends on the size of the gain or loss. She then opens a new trading position by purchasing one share of another stock.

The investor's objective is to time the repeated sales and purchases to maximize her average realization utility per period. Due to the simple nature of this problem, the optimal strategy is to pick two fixed values, $L < 0 < G$, and sell the first time the stock price increases by G dollars or decreases by L dollars from the original purchase price. For a given strategy, G and L, the average realization utility per period is

$$\frac{P(G,L) \cdot u(G) + [1 - P(G,L)] \cdot u(L)}{T(G,L)}, \tag{1}$$

where P is the probability that the gain is ultimately realized and T is the average duration of each holding period. Because each investment episode is identical ex ante, the sequence of investments is a renewal process. Equation 1 is a statement of the Elementary Renewal Theorem. The function $u(\cdot)$ measures the amount of realization utility that the investor receives depending on the size of the gain or loss. We pick the curvature of this function to capture the property of diminishing sensitivity discussed earlier.

The two functions, $P(G, L)$ and $T(G, L)$, can be determined by iterated expectations and are

$$P(G, L) = \frac{1 - \beta^L}{\beta^G - \beta^L} \qquad T(G, L) = \frac{(\beta^L - 1) \cdot G - (\beta^G - 1) \cdot L}{(1 - 2\pi) \cdot (\beta^G - \beta^L)},$$

where $\beta = (1 - \pi)/\pi$. The $u(\cdot)$ function we use for measuring realization utility is a special case of the Tversky-Kahneman (JRU 1992) utility function

$$u(x) = \begin{cases} (x)^\alpha & x \geq 0 \\ -(-x)^\alpha & x < 0 \end{cases},$$

where x is the size of the gain or loss, and α $(0 \leq \alpha \leq 1)$ is a parameter that measures the degree of diminishing sensitivity. For $\alpha = 1$ sensitivity does not diminish at all; the smaller is α the more quickly does sensitivity diminish. Table 1 shows the average realization utility per period for two different specifications.

Table 1: Average realization utility per period

Panel A: Utility parameter: $\alpha = 1$

			L			
G	−1	−2	−3	−4	−5	−∞
1	0.20	0.20	0.20	0.20	0.20	0.20
2	0.20	0.20	0.20	0.20	0.20	0.20
3	0.20	0.20	0.20	0.20	0.20	0.20
4	0.20	0.20	0.20	0.20	0.20	0.20

Panel B: Utility parameter: $\alpha = 0.5$

			L			
G	−1	−2	−3	−4	−5	−∞
1	0.20	**0.27**	0.26	0.25	0.24	0.20
2	0.07	0.14	0.16	0.16	0.16	0.14
3	0.04	0.10	0.12	0.12	0.12	0.12
4	0.03	0.08	0.09	0.10	0.10	0.10

The per-period probability of a price increase (π) is 60%. The optimal gain-loss realization strategy is indicated in bold.

Take good news in small doses and bad news in large ones

In Panel A, the utility parameter α is one. In this case, realization utility is linear in the size of gain or loss; that is, there is no diminishing sensitivity. This means that the marginal benefit of a one dollar gain is always the same regardless of how large a gain or loss is. As a result, all strategies yield the same average realization utility per period of 0.20, which is just the average per-period increase in the stock price.

In Panel B, the utility parameter α is 0.5. In this case, the first dollar gained or lost changes utility by one while the tenth dollar of a gain or loss changes utility by only 0.162, so there is a strong degree of diminishing sensitivity. Now there is a unique optimal trading strategy; in every trading episode, the investor waits until the stock price goes up by one dollar or goes down by two dollars before closing it out and opening a new position. The investor is willing to realize losses in this case because a two dollar loss yields a total decrease of 1.4 units of realization utility, whereas two separate one-dollar gains have a total benefit of 2 units of realization utility. Although closing out an underwater position is immediately painful, it frees up cash for future investments that on average generate trading profits and more utility in the future. (Specifically, keeping G = 1 and moving L from –3 to –2 increases the probability of realizing a loss from 12.3% to 21.1%, but it also shortens the average duration of each trading episode from 2.54 to 1.84 periods.) The property of diminishing sensitivity makes a sequence of small gains more beneficial to the investor than an occasional large loss. Together with realization utility, this generates the prediction that the investor will optimally realize small frequent gains and large infrequent losses.

This simple illustration highlights the results of our model. In the original RFS paper, we also consider many other factors. We examine the more realistic case of proportional rather than absolute price movements. We do not restrict purchases to a single share so the size of the investment affects investors' trading decisions. We consider transaction costs and show that their presence makes investors defer sales until both gains and losses are larger because trading frequently is very costly. Finally, we incorporate positive time discounting, the notion that getting a gain now means more to the investor than getting the same gain a year from now. This accelerates the realization of gains but can either retard or hasten the taking of losses. On the one hand, the investor naturally wants to postpone painful loss taking; on the other hand, the investor has an incentive to accelerate loss taking because by doing so she can realize future gains sooner.

Our model sheds light on a number of puzzling empirical facts. Among these, the most obvious one is something called the *disposition effect*. This is an empirically robust pattern that investors have higher propensities to sell stocks that have risen in value than stocks that have fallen in value. This is puzzling because empirical research shows that stocks display momentum: stocks that have done well tend to continue to do well in the future, while stocks that have done poorly tend to do poorly in the future. So, if investors do trade based on information, they should exhibit the opposite of the disposition effect. (Odean (JF 1998) examines other explanations such as portfolio rebalancing and tax motives and finds that these cannot explain the disposition effect.)

Under the framework of realization utility with diminishing sensitivity, however, the disposition effect naturally arises. On the one hand, investors tend to sell stocks that have risen in value because by doing so they get positive feelings (positive realization utility). On the other hand, even though selling stock at a loss brings investors immediate negative feelings, closing out losing positions allows investors to take on new investments that on average accelerate future gains and the good feelings when realizing them. It is only when investors have diminishing sensitivity that the good feelings from realizing frequent small gains can more than offset the bad feelings from realizing infrequent large losses.

In our paper, we provide some detailed calibration exercises. Using reasonable parameter values, our model predictions match the magnitudes and frequencies of realized gains and losses and the frequencies of paper gains and losses as observed in the trading data of Odean (JF 1998) and Dhar-Zhu (MS 2006).

Volatility and trading volume

Another empirical pattern that is consistent with our model is the flattening of the security market line. Ang-Hodrick-Xing-Zhang (JF 2006) document that high-beta and high-residual-risk stocks have smaller expected returns than predicted by equilibrium models such as the capital asset pricing model. Our framework gives a novel explanation for this. Stocks with higher volatilities, whether systematic or idiosyncratic, provide more opportunities for investors to earn realization-utility benefits. As a result, these investors tend to hold more of the highly volatile stocks than predicted by other equilibrium models. The excess demand of realization utility investors pushes up the prices of these stocks and decreases their expected returns.

Other findings that our model can help explain include the observations that individual investors trade excessively despite their underperforming market benchmarks even before transaction costs, that the trading volume is higher in rising markets than it is in falling markets, that investors have a higher propensity to sell a stock once its price rises above its recent historical high, that highly valued assets are heavily traded, that there is an empirical V-shaped pattern between the probability of selling a stock and its unrealized paper gain, and that investors hold and trade individual stocks and do not diversify their portfolios as much possible.

Finally, our model generates some new testable predictions particularly if the reference level, the benchmark with which investors divide their trading gains from losses, depends on the history of stock prices. One prediction is that total volume will be higher in a market that rises or falls quickly followed by a slowly trending period than in a market with the opposite pattern. And in both of these markets, the volume should exceed that in a market with a slow steady rise or fall of the same total magnitude.

In summary, our model highlights the important roles that realization utility and diminishing sensitivity play in understanding investors' trading behaviors. It helps explain of the empirical facts that previously seemed puzzling. We hope that future research can provide additional qualitative and quantitative tests of this framework and further compare our model with alternative theories.

Marinus van Reymerswaele (1490–1546): Money-changers. Flemish, 1548. Reymerswaele is known for his paintings documenting the flourishing economic activity of northern Europe in the 16th century. Does this painting show the age old theme of ridiculing the supposed greed of the affluent money changers or bankers?

Yan Li and Liyan Yang

Prospect theory, the disposition effect, and asset prices

One of the most studied individual trading behaviors is the "disposition effect": investors have a greater tendency to sell assets that have risen in value since purchase than those that have fallen. Because none of the most obvious rational explanations, such as portfolio rebalancing or information story, can entirely account for the disposition effect, an alternative view based on prospect theory has gained favor.

Prospect theory has several salient features: (i) investors evaluate outcomes, not according to final wealth levels, but according to their perception of gains and losses relative to a reference point, typically the purchase price; (ii) investors are more sensitive to losses than to gains of the same magnitude (loss aversion); and (iii) investors are risk-averse for gains and risk-seeking for losses (diminishing sensitivity).

What if everyone acts this way?

It is the diminishing sensitivity feature of prospect theory that researchers often cite as the underlying mechanism of the disposition effect: if a stock is traded at a gain, then the investor is in his risk-averse domain and is inclined to sell the stock; if a stock is traded at a loss, then the investor is in his risk-seeking domain and tends to hold on to the stock. Recent theoretical models, however, suggest that the link from prospect theory and the disposition effect is more nuanced. Specifically, prospect theory will fail to predict the disposition effect when the expected return is high (for an investor to buy a stock, its expected return must be reasonably high, which might encourage the investor to take more risk to continue to hold the stock after a gain), or when returns are positively correlated over time (after a gain, the investor expects another increase in price and would be more likely to hold on to the stock). In all these existing models, stock returns are assumed to follow an exogenous process, and it remains unclear whether prospect theory still predicts a disposition effect when stock returns are affected by the trading of prospect-theory investors.

This question is particularly important in light of two other related asset pricing literatures. First, the loss aversion feature of prospect theory has been used to explain the historical high equity premium. Second, the disposition effect has been employed to generate price momentum, because the disposition effect creates a wedge between a stock's fundamental value and its equilibrium

price, leading to price underreaction to information. However, in principle, both a high equity premium and price momentum will make prospect theory less likely to deliver the disposition effect, as suggested by the above-mentioned partial equilibrium models. To examine whether and to what extent prospect theory can simultaneously explain the disposition effect, the momentum effect, and the equity premium puzzle (which are three of the most well-known puzzles and which have been separately investigated using prospect theory), and in particular, to give a definitive answer to whether prospect theory can generate the disposition effect, we need a general equilibrium model in which prospect-theory investors trade stocks and affect stock prices.

Our general equilibrium model

In our JFE paper, we build such a model to examine the implications of prospect theory for trading, pricing, and volumes. We adopt an overlapping-generation (OLG) setup with three generations: age-1, age-2, and age-3. Our OLG setup can be understood as a stylized way of describing how different types of investors in real markets interact with each other. The age-1 investors correspond to new participants to the market; the age-2 investors correspond to the discretionary traders who have been sitting in the market for some time; and the age-3 investors correspond to pure noise investors. All investors trade a risky asset (stock) and a risk-free asset (bond). In each period t, the bond is traded at the constant gross risk-free rate $R_f > 1$, and the stock's price P_t is determined by investors' trading behaviors.

The stock is a claim to a stream of dividends. The dividend growth rate is independently and identically distributed (i.i.d.) over time, and it can equally likely take either a high value θ_H or a low value θ_L (with $0 < \theta_L < \theta_H$). However, some investors are more optimistic than others in the sense that they believe the next period dividend growth rate is more likely to be high. This heterogeneity implies that in each period those optimistic investors will hold the stock in equilibrium. Also, for a given investor, he may experience belief changes: it is possible that he was initially optimistic and then becomes pessimistic, and so, he may first buy the stock and later want to sell it, which creates the possibility for prospect theory to affect his selling behaviors.

When investor i enters the market, he is endowed with $W_{1,i}$ units of consumption good. He can trade at ages 1 and 2, leaving his final wealth as $W_{3,i}$ and his capital gains/losses as $X_{3,i}$. His time t utility, U_t^i, is then given by

$$U_t^i = E_t^i[v(X_{3,i})],$$

where $E_t^i[\cdot]$ is his expectation at time t, and

$$X_{3,i} = W_{3,i} - R_f^2 \cdot W_{1,i}$$

measures the capital gains/losses, and where

$$v(x) = \begin{cases} x^\alpha, & \text{if } x \geq 0 \\ -\lambda \cdot (-x)^\alpha, & \text{if } x < 0 \end{cases}$$

is the standard prospect-theory value function that is used to evaluate the gains/losses.

Parameter $\alpha \in (0, 1]$ governs the value function's concavity/convexity and parameter $\lambda \in [1, \infty)$ controls loss aversion. When $\alpha = \lambda = 1$, both diminishing sensitivity and loss aversion vanish, reducing the preferences to a standard risk-neutral utility representation. Figure 1 plots v(x) for the cases of $\alpha = 0.5$ and $\alpha = 2.25$: it is concave over gains and convex over losses, and it has a kink at the origin.

Figure 1: The value function of prospect theory
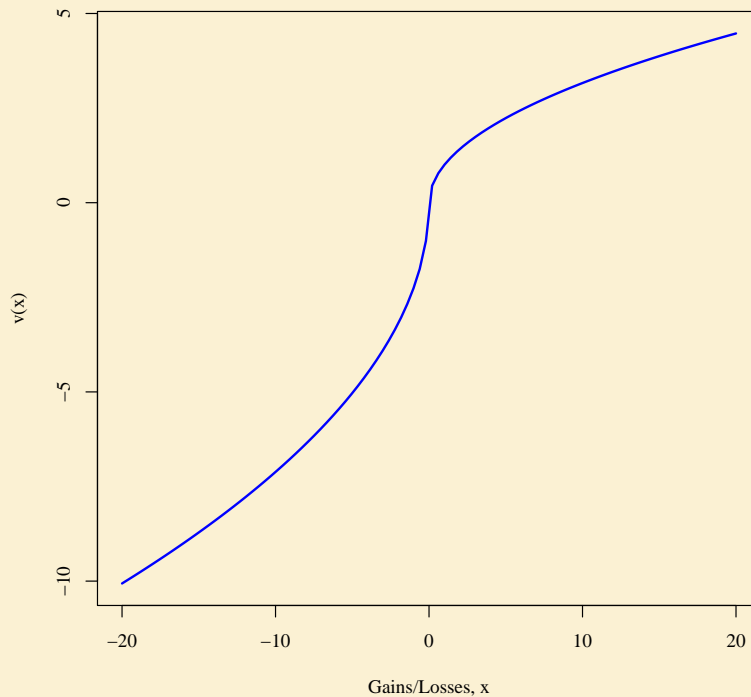
The parameters to the v(x) function here are $\alpha = 0.5$ and $\lambda = 2.25$.

Our financial market works as follows. In each period, before trading occurs, the stocks are initially held by the age-2 and age-3 investors who were initially optimistic and already purchased stocks in previous periods. All those age-3 investors sell stocks because they will exit the economy at the end of the period. Whether age-2 investors continue to hold the stock depends on their expectations of future dividend growth rate, and the sufficient pessimistic ones end up selling the stock. The (reversed) disposition effect concerns the different behaviors of those age-2 investors as a group in good versus bad dividend news. Their state-dependent behaviors will influence prices by shifting the aggregate demand function, generating momentum or reversal. The age-1 investors who just enter the market and the remaining age-2 investors who didn't purchase the stock last

period decide whether to buy the stock by comparing utilities from buying with those from not buying; those optimistic among them will end up buying the stock.

We analytically solve the equilibrium in the case of risk-neutral utility ($\alpha = \lambda = 1$), which serves as our benchmark economy. In the case of prospect-theory utility, we use numerical method to solve the equilibrium. In the numerical analysis, we take one period to be six months, and set $\theta_H = 1.19$ and $\theta_L = 0.83$, to match the empirical values of the mean and volatility of the net annual dividend growth rate. The risk-free rate is set at $R_f = 1.0191$.

Table 1: Trading and pricing implications of prospect theory

Panel A: Implications of diminishing sensitivity (by varying α and fixing $\lambda = 1$)

Variables	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.88$	$\alpha = 1$
DispEffect	1.93	1.61	1.13	1.00
MomEffect, in %	5.62	3.28	0.76	0.00
EqPremium, in %	1.57	1.54	0.93	0.00

Panel B: Implications of loss aversion (by varying λ and fixing $\alpha = 1$)

Variables	$\lambda = 1$	$\lambda = 2.25$	$\lambda = 2$	$\lambda = 4$
DispEffect	1.00	0.95	0.89	0.81
MomEffect, in %	0.00	−0.25	−0.56	−1.03
EqPremium, in %	0.00	3.77	5.61	7.91

Panel C: Quantitative analysis (by varying α and fixing $\lambda = 2.25$)

Variables	$\alpha = 0.37$	$\alpha = 0.48$	$\alpha = 0.52$	$\alpha = 0.61$	$\alpha = 0.88$	Empirical
DispEffect	2.15	1.79	1.68	1.49	1.07	2.24
MomEffect, in %	4.97	3.59	3.16	2.32	0.34	5.27
EqPremium, in %	5.09	5.08	5.02	4.99	4.63	3.84

This table uses simulations to examine the implications of prospect theory for the disposition effect, the momentum effect, and the equity premium. In the simulation, we take one period to be six months. In each period, a good dividend shock and a poor dividend shock are equally likely. Parameters θ_H and θ_L are calibrated at 1.1913 and 0.8310, to match the mean and volatility of the annualized dividend growth rate. The risk-free rate is set at $R_f = 1.0191$. The empirical values in Panel C are borrowed from previous studies or are computed based on NYSE/Amex data from 1926–2009. Parameter α determines diminishing sensitivity, and parameter λ controls loss aversion. When $\alpha < 1$, diminishing sensitivity is active, and a smaller α means that investors are more risk averse over gains and more risk loving over losses. When $\lambda > 1$, loss aversion is active, and a larger λ means that investors are more loss averse.

We report the results in Table 1. The variable DispEffect is the ratio of "proportion of gains realized" (PGR) to "proportion of losses realized" (PLR). If DispEffect > 1, then investors exhibit a disposition effect in our model, and if DispEffect < 1, they exhibit a reversed disposition effect. We measure momentum as $MomEffect = E(R_{t+1} | \theta_t = \theta_H) - E(R_{t+1} | \theta_t = \theta_L)$, where R_{t+1} is the gross return on the stock between time t and t + 1. That is, MomEffect is the difference in the expected return following positive dividend news and following negative dividend news. If MomEffect > 0, then there is momentum in stock returns, and if MomEffect < 0, then there is reversal. The variable $EqPremium = E(R_t - R_f)$ is the equity premium, i.e., the average stock return in excess of the risk-free rate.

Diminishing sensitivity, momentum, and the equity premium

Panel A examines the implications of diminishing sensitivity by conducting comparative static analysis with respect to α. We here set parameter λ at 1 to remove the loss aversion feature of the preference. In particular, when $\alpha = 1$ (i.e., when investors are risk-neutral), investors do not exhibit the (reversed) disposition effect, and returns do not exhibit momentum(reversal) and have a mean equal to the risk-free rate. However, as long as $\alpha < 1$, the diminishing sensitivity feature of prospect theory drives a disposition effect, a momentum effect, and a positive equity premium.

The intuition is as follows. When a stock experiences good news and increases in value relative to its purchase price, investors who previously purchased it are in the concave, risk-averse region of the value function of prospect theory. Conversely, when a stock experiences bad news, its investors face capital losses, and they are in the convex, risk-seeking region. So, facing good news, they are keen to sell the stock (i.e., a disposition effect), and their selling makes the stock price underreact to the initial good news in equilibrium, leading to subsequent higher returns (i.e., a momentum effect). Facing bad news, they are reluctant to sell, absent a premium; the price underreacts to the initial bad news, giving rise to subsequent lower returns. The intuition for the positive equity premium is subtle and it is driven by the behavior of age-1 investors who are more likely to wait to buy the stock in the future when their value function is more curved.

Loss aversion, reverse disposition, and price reversal

Panel B examines the implications of loss aversion by conducting comparative static analysis with respect to the parameter λ and by setting parameter $\alpha = 1$ to remove the diminishing sensitivity feature. We find that, loss aversion drives a reversed disposition effect, a reversal in stock returns, and a positive equity premium. The intuition is the following. Loss aversion means that the prospect-theory value function has a kink at the origin. Investors are afraid of holding stocks if they are close to the kink. As is well known in the literature, loss aversion produces positive equity premiums in equilibrium, and thus, good news will push investors far from the kink and bad news will push them close to the kink. As a result, when facing gains, investors are more likely to hold stocks (i.e., a reversed disposition effect); the increased demand resulting from the reversed disposition effect causes the stock price to "overreact" to the initial good news, pushing the current

price even higher and leading to lower subsequent stock returns (i.e., a reversal in stock returns). This effect is symmetric, so that when a stock experiences bad news, the opposite happens.

Diminishing sensitivity is more important than loss aversion

Given that different components of prospect theory often make opposite predictions regarding investors' trading and asset prices, it is nontrivial to examine whether prospect theory can explain the data in economies with preference parameters α and λ set at their empirical values. Panel C conducts such a quantitative analysis. Specifically, abundant empirical studies estimate λ to be close to 2, and hence we fix λ at 2.25. Previous empirical studies have obtained different estimates for the value of α, and we report results for all the possible estimates of α: 0.37, 0.48, 0.52, 0.61, and 0.88. We also report in Panel C the historical values for the variables of interest, which are either borrowed from previous studies or computed based on NYSE/Amex data from 1926–2009. We find that, for all five possible values of α, the diminishing sensitivity component of prospect theory dominates the loss aversion component, which implies that prospect theory indeed helps to explain the disposition effect, the momentum effect, and the high equity premium. In particular, for the case of $\alpha = 0.37$, the model-generated variables are quite close to their empirical values.

In our JFE paper, we also use our model to explore the volume and pricing implications of dividend volatility and skewness and suggest new testable empirical predictions. Our analysis highlights the importance of using prospect theory to advance our understanding of individual trading behavior and salient asset-pricing phenomena.

Peter Paul Ruben: Tribute Money. Flemish Baroque, 1612. Jesus advises the shocked Pharisees to do the right thing by paying taxes to Caesar. Tribute Money became a common theme in the 16th century, because of the conflict between the Catholic Church and the Roman Emperor Charles V.

Alessandro Beber and Marco Pagano

Short-selling bans around the world: lessons from the financial crisis

On 19 September 2008—just as the failure of Lehman Brothers had shaken investors' confidence in banks' solvency and sent stocks into free fall—the U.S. Securities and Exchange Commission (SEC) prohibited "short sales" of financial companies' stocks. The hope was that this would stem the tide of sales and help support bank stock prices.

The SEC's move sparked worldwide herding by regulators. In the subsequent weeks and months, most stock exchange regulators around the globe issued bans or regulatory constraints on short selling. Some of the bans were "naked," i.e. only ruled out sales where the seller does not borrow the stock in time to deliver it to the buyer within the standard settlement period (naked short sales). Other bans were "covered," ruling out also sales where the seller manages to borrow the stock (covered short sales). By the end of October 2008, no less than 20 countries around the globe had imposed some form of short-selling ban.

These hurried interventions, which varied considerably in intensity, scope and duration, were invariably presented as measures to restore the orderly functioning of security markets and limit unwarranted drops in securities prices. Did they achieve their stated purpose? And did they have any negative side effects?

Both theoretical arguments and previous evidence would have advised greater caution. The effectiveness of short-selling bans in supporting stock prices is controversial, and several previous studies alerted that such bans can damage stock market liquidity and slow down the speed at which new information is impounded in stock prices. Because the crisis was accompanied by a widespread and steep increase in bid-ask spreads in stock markets, it is important to understand whether, and to what extent, short-selling bans contributed to their increase, and therefore reduced stock market liquidity.

Was the SEC right?

Now that economists have canvassed the evidence about the short-selling bans during the crisis, it is possible to evaluate this policy intervention. Boehmer-Jones-Zhang (WP 2009) have analyzed the response of liquidity measures to the short-selling ban imposed by the SEC from September 19 to October 8, exploiting the difference between the financial sector stocks targeted by the

ban and those that were not. They have found that liquidity—as measured by spreads and price impacts—deteriorated significantly for stocks subject to the ban.

But did the SEC at least manage to achieve its stated objective, that is, stem the collapse of stock prices? Even this is unclear. Boehmer-Jones-Zhang (WP 2009) document large price increases for banned stocks upon announcement of the ban, followed by gradual decreases during the ban period. Yet they recognize that the correlation with the ban could be spurious, as the prices of U.S. financial stocks could have been affected by the accompanying announcement of the U.S. bank bail-out program—the Troubled Asset Relief Program (TARP). Their skepticism is reinforced by the finding that stocks that were later added to the ban list experienced no positive share price effects. However, Harris-Namvar-Phillips (WP 2009) try to control for the TARP legislation and find that the positive abnormal returns for banned stocks cannot be explained by a TARP fund index.

Clearly, reliance on data from the U.S.—where the start of the short-selling ban on financials coincided with bank bailout announcements—makes it hard to identify the price effects of the ban. International evidence can be valuable in this respect, because short-selling bans in several other countries were not accompanied by bank bailout announcements. Moreover, in many countries bans also applied to non-financial stocks, and in other countries financial stocks were simply not banned.

New worldwide evidence

In Beber-Pagano (JF 2013), we harness the large amount of evidence that short-selling bans generated during the crisis, assembling daily data for nearly 17,000 stocks from 30 countries, for the period spanning from January 2008 to June 2009. A key feature of the data is that short-selling restrictions were imposed and lifted at different dates in different countries. They often applied to different sets of stocks (only financials in some countries, all stocks in others) and featured different degrees of stringency. This variation in short-selling regimes is important because it makes the data ideally suited to identify the effects of the bans through panel data techniques. The extent of variation in short-selling regimes between September 2008 and June 2009 is illustrated in Figure 1 via color-coded lines. Dark and light blue lines correspond to naked bans of financial and non-financial stocks, respectively. Red lines indicate covered bans for financial stocks, while orange lines correspond to covered bans of non-financial stocks. The figure illustrates the variety of regimes and regime durations across countries, as well as the complex regime variation over time, even within the same country (the extreme example here being Italy).

Figure 1: Short-selling ban regimes around the world, Sep08 To Jun09
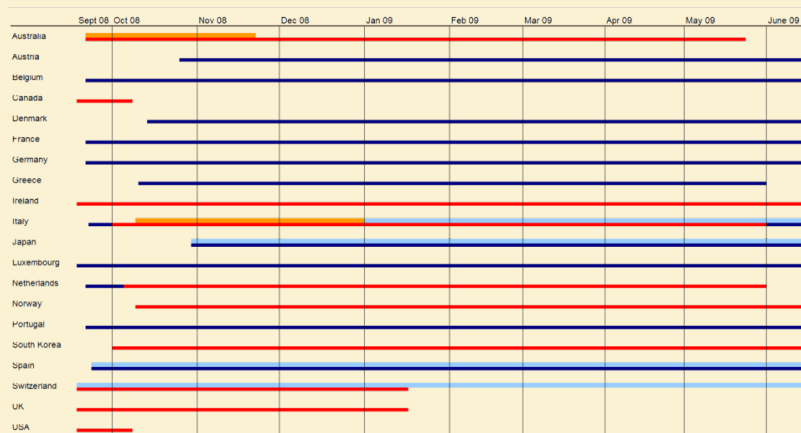
A visual comparison across different countries.

In our empirical analysis we study whether stocks that were subject to short-selling restrictions featured different price performance, liquidity or informational efficiency when benchmarked against stocks exempt from such restrictions. In performing this comparison, we control for time-invariant stock characteristics, as well as for return volatility and for common risk factors. The latter controls are important, because during the crisis increased uncertainty and acute funding problems were likely to have affected stock market liquidity throughout the world.

Figure 2: Cumulative abnormal returns in the U.S. for stocks subject to covered bans and for exempt stocks
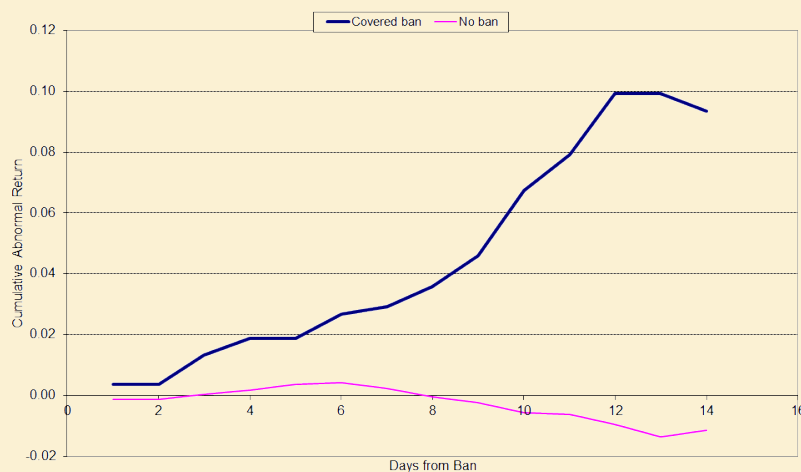
The figure plots cumulative abnormal returns in the 14 trading days after the ban date, which is date 0 in the graph.

Our results indicate that the bans have not been associated with better stock price performance, the U.S. being the exception. The most immediate evidence on this point is obtained by comparing

post-ban median cumulative excess returns—with respect to market indices—for stocks subject to bans with those of exempt stocks. Figure 2 shows that the median cumulative excess return of U.S. financial stocks, which were subjected to a covered ban, exceeded that of exempt stocks throughout the 14 trading days after the ban inception (date 0 in the figure), a finding that agrees with that reported by Boehmer-Jones-Zhang (WP 2009). But Figure 3 shows that this is not the case for the other countries in our sample: the line corresponding to the median excess return on stocks subject to naked and covered bans is very close to that for exempt stocks, and it lies above it only in about half of the first 60 days of trading after the inception of the ban. Because the confounding factor of simultaneous policy measures is likely to convey a more accurate picture of the effects of short-selling bans on stock returns than the evidence for the U.S. shown in Figure 2.

Figure 3: Cumulative abnormal returns in countries with partial bans (except the U.S.) for stocks subject to ban and exempt stocks
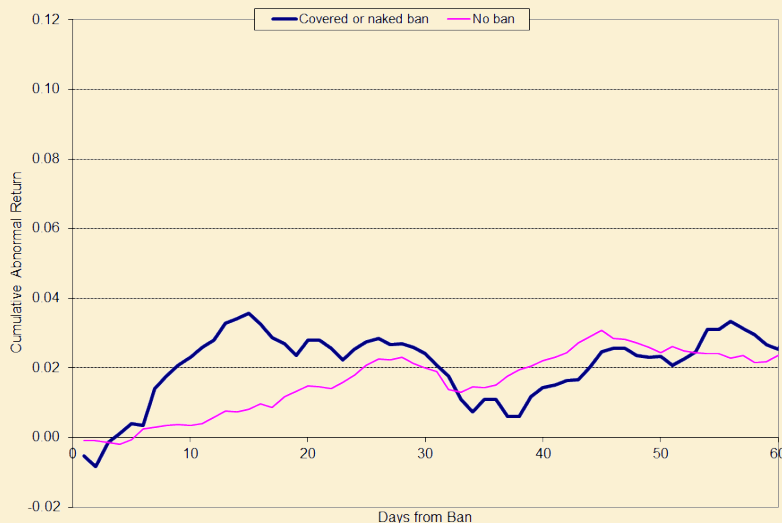
The figure plots cumulative abnormal returns in the 60 trading days after the ban date, which is date 0 in the graph.

This conclusion is confirmed and actually reinforced by the econometric analysis. When we use our entire data set, bans on covered short sales turn out to be correlated with significantly lower excess returns relative to stocks unaffected by the ban, while bans on naked sales and disclosure obligation do not have a significant correlation with excess returns. When we consider countries with short-selling bans on financials only, bans turn out to be correlated with positive excess returns only for the U.S., not for other countries. But, as noted above, the positive correlation for the US may be spurious.

Hence, in contrast to the regulators' hopes, worldwide evidence indicates that short-selling bans have at best left stock prices unaffected, and at worst may have contributed to their decline.

Moreover, we find that short-selling bans imposed during the crisis had unintended but important negative consequences on liquidity and price discovery. Our results indicate that the bans are associated with a statistically and economically significant increase in bid-ask spreads throughout the world. In contrast, the obligation to disclose short sales is associated with a significant decrease in bid-ask spreads.

Figure 4: Average bid-ask spread of stocks subject to bans and of matched exempt stocks for countries with partial bans.
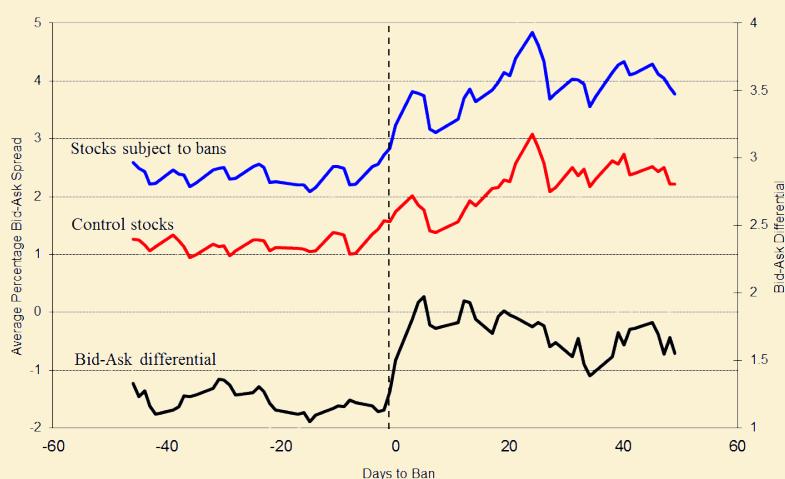
The lines plots the three-day moving average of the bid-ask spread's cross-sectional average for stocks subject to bans and control stocks (left scale) and their differential (right scale), in a 50-day window around the ban inception date (date 0). The data correspond to countries with partial bans: Belgium, Canada, Germany, Denmark, France, the Netherlands, Ireland, Norway, Austria, Portugal, the U.K., and the U.S.

In Figure 4 we plot the average bid-ask spreads of stocks subject to the ban and their matching stocks during our event window, as well as that of their differential. The matching stock is the exempt stock traded in the same country and with the same option listing status that is closest in terms of market capitalization and stock price level. The figure shows that immediately after the ban date the gap between the average bid-ask spread of banned stocks and that of exempt stocks widens, suggesting that the ban had a detrimental effect on liquidity. This conclusion is again confirmed by the econometric analysis.

We also find evidence that short-selling bans made stock returns more correlated with their own past values, that is, slowed down the reaction of stock prices to new information. This slowdown in price discovery especially when negative news are concerned, is in line with the findings of previous empirical studies, for instance Bris-Goetzmann-Zhu (JF 2007), and with the predictions of the theory. By restraining the trading activity of informed traders with negative information about fundamentals, a short-selling ban slows down the speed at which news are impounded in market prices, and more so in bear market phases.

Lesson learned?

To sum up, the evidence suggests that the knee-jerk reaction of most stock exchange regulators to the financial crisis—imposing bans or regulatory constraints on short-selling—was at best neutral in its effects on stock prices. The impact on market liquidity was clearly detrimental. Moreover, the bans were associated with slower price discovery.

Perhaps the main social payoff of this worldwide policy experiment has been that of generating a large amount of evidence about the effects of short-selling bans. The conclusion suggested by this evidence is best summarized by the words of the former SEC Chairman Christopher Cox on 31 December 2008: "Knowing what we know now, [we] would not do it again. The costs appear to outweigh the benefits." Unfortunately, in sharp contrast with this assessment and with our strong evidence, security market regulators in some European countries have reacted to the more recent crisis events by imposing again short-selling bans on financial stocks. Apparently the lesson was not learned, at least by some regulators.

Titian (1488–1576): Diana and Castillo. Italy, 1559.
In 2008, the Duke of Sutherland announced that he would be willing to sell two Titian masterpieces (the other being Diana and Actaeon, shown after the next article) much below market value—for about $150m, half its estimated true value—to the UK National Art Galleries. After they declined, he decided to auction them off. After frantic fund-raising and much criticism for government funding in hard economic times, the UK finally managed to buy the paintings, after all. The United Kingdom thus barely avoided to what would have been the equivalent of Michelangelo's David leaving Italy. But...wasn't Tiziano Italian, anyway?

A number of trends over the two decades leading to the financial crisis that started in 2007 made it much easier for homeowners to refinance their mortgages to take advantage of declining interest rates, increasing housing prices, or both. In Khandani-Lo-Merton (JFE 2013), we argue that during periods of rising home prices, falling interest rates, and increasingly competitive and efficient refinancing markets, cash-out refinancing acts like a ratchet: homeowner leverage increases incrementally as real-estate values appreciate, but cannot decrease incrementally as real-estate values decline. This self-synchronizing "ratchet effect" can create significant systemic risk in an otherwise geographically and temporally diverse pool of mortgages. We show that even in the absence of any dysfunctional behavior such as excessive risk-taking, fraud, regulatory forbearance, political intervention, and predatory borrowing and lending, large system-wide shocks can occur in the housing and mortgage markets. The mechanism behind this systemic risk is subtle and complex, arising from the confluence of three familiar and individually welfare-improving economic trends.

Our approach

To gauge the magnitude of the potential risk caused by the refinancing ratchet effect, we created a numerical simulation of the U.S. housing and mortgage markets that matches the size and growth of this market over the last few decades using data from 1919, because over 93% of homes in use today were built after 1919.

We modeled realistic heterogeneity in our simulations by incorporating geographical diversity in the rate of home price appreciation, diversity in initial purchase price of homes, as well as diversity in the types and size of mortgages used to purchase each home based on several data sources such as the 2007 American Housing Survey, the Federal Housing Finance Agency, and the Case-Shiller housing price indices. By following mortgages associated with each home and by specifying reasonable behavioral rules for the typical homeowner's equity extraction decision, we were able to match some of the major trends in this market over the past several decades. For example, as shown in Figure 1, our simulations can match the dramatic rise in outstanding total mortgages and cumulative equity extractions in the last two decades. (Cash-out refinancing takes place according to the base-case specification outlined in the JFE paper.)

Figure 1: Simulated and actual mortgage debt outstanding and cumulative equity extractions
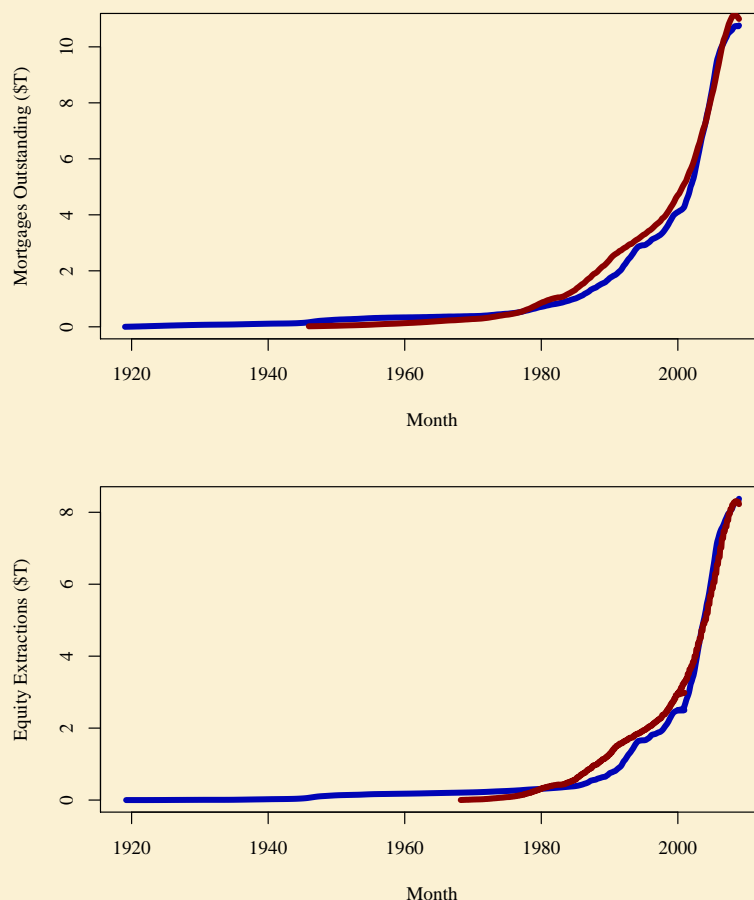
In the top figure, the blue line are outstanding mortgages simulated under the calibrated uniform rule. The maroon line are the actual total mortgages outstanding from the Federal Reserve Flow-of-Funds Accounts. In the bottom figure, the blue line are cumulative equity extractions simulated under the calibrated uniform rule. The maroon line are the total mortgage liability series from Greenspan-Kennedy (OREP 2008).

Option-based evaluation of losses and risks

Armed with a properly calibrated simulation of the U.S. residential mortgage market, we turn to assessing the systemic risk posed by the refinancing ratchet effect. Given our assumption that all mortgages in our simulations are non-recourse loans—collateralized only by the value of the underlying real estate—the homeowner has a guarantee, or put option, that allows him to put or "sell" the home to the lender at the remaining value of the loan if the value of the home declines to below the outstanding mortgage.

As mortgages are placed in various structured products like collateralized mortgage obligations and then sold and re-sold to banks, asset management firms, or government-sponsored enterprises (GSEs), the ultimate entities exposed to these guarantees may be masked. However, it is clear that all mortgage lenders must, in the aggregate, be holding the guarantees provided to all homeowners. To the extent that some owners may be liable for the deficiency in their collateral value through recourse, those owners share some of the burden of the loss caused by a decline in home prices. Therefore the estimated economic loss and various risk metrics should be viewed as the amount of economic loss or risk exposure for the lenders and the subset of borrowers that are legally held responsible via recourse in their mortgages.

Table 1 shows that the ratchet effect alone is capable of generating the magnitude of losses suffered by mortgage lenders during the Financial Crisis of 2007–2009, yielding $1.7 trillion in losses for mortgage-lending institutions since June 2006 compared to $330 billion in the case of no-equity extractions.

A benefit of using option-pricing technology for this calculation is that in addition to giving us a numerical value for the total economic loss, it provides us with general forward-looking risk measures based on the sensitivities of the value of mortgages' embedded put options to changes in the level or volatility of home prices. For example, as reported in Table 1, in the first quarter of 2005 we estimate that the aggregate value of all embedded put options would increase by $18.17 billion for each 1% drop in home prices. By the last quarter of 2008, this sensitivity almost doubled to $38.13 billion for each 1% drop in home values compared to only $7.42 billion in which no equity had been extracted. This increase is due to the large convexity, or gamma using option-based measures, in the value of embedded options. As reported in Table 1, we estimate that gamma was $573.79 million per 1% drop in home values in the first quarter of 2005, which increased to $801.13 million for each 1% drop in home values by the last quarter of 2008.

The size and increase in the gammas of these options indicate substantial non-linearity in the risk of the mortgage system that may need to be accounted for in systemic risk measurement and analysis. We estimate that the total value of the embedded put options in non-recourse mortgages would increase by approximately $70 to $80 billion for each 1% increase in home price volatility in the years leading up to the crisis—known as the options' "vega" —compared to only $10 to $15 billion for each 1% increase in home price volatility if no equity had been extracted.

Table 1: The ratchet effect

| | No Cash-Out | | | | Cash-Out | | | |
	Put Value	Delta	Gamma	Vega	Put Value	Delta	Gamma	Vega
Mar-05	64.90	2.2	79	11	566.60	18	574	75
Jun-05	65.70	2.3	80	11	568.80	18	581	76
Sep-05	69.20	2.4	83	11	592.30	19	598	78
Dec-05	74.20	2.5	87	12	601.30	19	611	80
Mar-06	81.50	2.7	91	12	621.10	20	625	81
Jun-06	87.10	2.9	97	13	611.70	20	631	81
Sep-06	100.90	3.2	105	13	644.90	21	648	82
Dec-06	114.90	3.6	113	14	698.70	22	673	83
Mar-07	126.80	3.9	120	15	748.40	23	696	84
Jun-07	137.30	4.1	127	15	782.80	24	715	85
Sep-07	153.30	4.5	136	16	831.20	25	735	85
Dec-07	178.90	5.1	146	16	951.00	27	771	86
Mar-08	221.80	5.8	156	16	1,185.10	31	813	84
Jun-08	252.40	6.4	161	16	1,345.20	34	829	82
Sep-08	278.40	6.8	164	16	1,465.40	35	824	79
Dec-08	330.20	7.4	165	16	1,727.20	38	801	74

Simulated time series of the aggregate value and sensitivities of total guarantees extended to homeowners by mortgage lenders for cash-out and no-cash-out refinancing scenarios for each quarter from 2005Q1 to 2008Q4 (put values are in $billions, deltas are in $billions per 1% decline in home prices, gammas are in $millions per 1% decline in home prices, and vegas are in $billions per 1% increase in home price volatility). Cash-out refinancing takes place according to the base-case specification outlined in the JFE paper.

Solution to the refinancing ratchet effect

Indivisibility and sole ownership of residential real-estate are two special characteristics of this asset class that make addressing the ratchet effect particularly challenging. Because the owner is typically the sole equity holder in an owner-occupied residential property, it is difficult to bring incrementally additional capital in to reduce risk by issuing new equity. Therefore, the only option available to homeowners in a declining market is to sell their homes, recognize their capital losses, and move into less expensive properties that satisfy their desired loan-to-value (LTV) ratio which imposes enormous financial and psychological costs.

A simple remedy is to require all mortgages to be recourse loans. If all mortgages were recourse loans and borrowers had uncorrelated sources of income, the additional income of the borrowers would create an extra level of protection for the lenders and, therefore, distribute the risk in the mortgage system between lenders and borrowers more evenly. However, the legal procedure for foreclosure and obtaining a deficiency judgment is complex, varying greatly from state to state. For

example, home mortgages are explicitly non-recourse in only 11 states. Even in certain populous states with recourse, such as Florida and Texas, generous homestead-exemption laws can make it virtually impossible for lenders to collect on deficiency judgments because borrowers can easily shield their assets.

For this reason, the losses and risks estimated using the methods discussed in this paper should be thought of as losses that can be passed to the borrowers to the extent that mortgage lenders have recourse and can obtain deficiency judgments against defaulting borrowers.

Conclusions

The fact that the refinancing ratchet effect arises only when three market conditions are simultaneously satisfied demonstrates that the recent financial crisis is subtle and may not be attributable to a single cause. Moreover, a number of the activities that gave rise to these three conditions are likely to be ones that we would not want to sharply curtail or outright ban because individually they are beneficial. While excessive risk-taking, overly aggressive lending practices, pro-cyclical regulations, and political pressures surely contributed to the recent problems in the U.S. housing market, our simulations show that even if all home-owners, lenders, investors, insurers, rating agencies, regulators, and policymakers behaved rationally, ethically, and with the purest of motives, financial crises could still occur. Therefore, we must acknowledge the possibility that no easy legislative or regulatory solutions may exist.

Titian (1488-1576). Diana and Actaeon. Italy, 1559. Purchased from the Duke of Sutherland by the U.K. National Galleries in 2009 for $79.1 million. This was the companion piece to Diana and the Castillo shown and described above.

Investor activism has become increasingly prevalent in Corporate America over the past two decades. Through acquiring equity ownership in underperforming firms with poor corporate governance practices, activist investors help the targeted firms improve performance through various forms of intervention. Over the last decade, a new breed of activist investors has been on the rise. This under-studied group of investors specializes in trading claims of distressed firms with an intention to influence reorganization, and hedge funds have emerged as the most active players in this field due to their ability to hold highly concentrated and illiquid positions as well as their minimum disclosure requirements. More specifically, these hedge fund investors aggressively acquire distressed debt claims and equity stakes of distressed companies, serve on the unsecured creditors committee or equity committee, and pursue the loan-to-own strategy, whereby a hedge fund acquires the debt of a distressed borrower with the intention of converting the acquired position into a controlling equity stake upon the firm's emergence from Chapter 11. In Jiang-Li--Wang (JF 2012), we study the roles of activist hedge funds in Chapter 11 firms and the effects of their presence on the nature and outcomes of the bankruptcy process.

Hedge fund involvement in largest 500 bankruptcies

To form our sample of study, we start started with the largest 500 U.S. Chapter 11 cases filed during the period from 1996 to 2007. We then obtain information on important milestones reached during restructuring (such as the extension of the exclusivity period, debtor-in-possession financing, approval of key employee retention plan (KERP), and top management turnover). Further, we gathered final outcomes such as emergence from bankruptcy, acquisition, or liquidation, from a variety of sources including BankruptcyData.com, Bankruptcy DataSource, PublicAccess to Court Electronic Records (PACER), and news searches in Factiva and LexisNexis. Next, we resort to BankruptcyData.com, 8K and 10K filings, proxy statements, Schedule 13D and Form 13F filings, and news searches to identify hedge-fund involvement in these distressed companies. We obtain firm-level financial and stock price information from Compustat, CapIQ, 10-K filings on EDGAR, and CRSP.

Table 1: Hedge-fund presence in Chapter 11 by timing and entry and their role

Panel A: Hedge-Fund Presence before bankruptcy

Hedge-Fund Presence	% of total 474 cases
Largest unsecured creditors	25.1
Largest shareholders	48.5
13D filing	7.0

Panel B: Hedge-Fund Presence during bankruptcy

Hedge-Fund Presence	% of total 474 cases
Unsecured creditors committee	38.5
Debtor-in-possession financing	9.1
13D filing	4.4
Equity committee	5.8

Panel C: Both before and during bankruptcy

Hedge-Fund presence	% of total 474 cases
Loan-to-own	27.7
Debt side	60.7
Equity side	53.4
Overall	87.4

Table 1 presents the overview of hedge-fund involvement during the Chapter 11 process, grouped by the timing of their entry and their roles, respectively. We show that 87% of the cases have publicly observable hedge-fund involvement in some form. Further, in 61% of the cases, hedge funds are present on the debt side (versus 53% in equity), and in 28% of the cases, hedge funds adopt a loan-to-own strategy. The five most active funds are: Oaktree Capital Management, Cerberus Capital Management, Loomis Sayles & Co, Appaloosa Management, and PPM America Special Investments Fund.

Hedge funds choose wisely and affect outcomes

The relationship between a hedge fund's presence and bankruptcy outcomes can be classified as one of two varieties:

1. a pure selection effect, whereby informed hedge funds merely pick the target that offers the best expected payoff, but do not affect the value of the underlying assets, and

2. a pure treatment effect, whereby hedge funds change the outcome and hence the value of the underlying assets even if they were randomly assigned to distressed firms.

A priori, a combination of these two effects is likely at work. Hedge funds are sophisticated investors that and could potentially profit from their company-picking skills even if they remain passive stakeholders. At the same time, hedge funds are likely to choose investment opportunities in which they can more effectively influence the outcome in their favor. It is worth noting that our measures for hedge-fund participation embed their activist roles. For example, if hedge funds can achieve the desired outcome just by picking the right companies without exerting influence during the Chapter 11 process, they could remain passive stakeholders without the costly voluntary effort of forming and serving on those committees.

To accommodate both the selection and the treatment effects, we use the following model:

$$\text{HFPart}_i^* = X_i \cdot \beta + \epsilon_i.$$

We set HFPart_i (unstarred) to 1 if $\text{HFPart}_i^* \geq 0$ and to 0 otherwise. The model is

$$\text{Outcome}_i = Z_i \cdot \gamma + \mu \cdot \text{HFPart}_i + \eta_i$$

HFPart is an indicator variable for hedge-fund participation in various ways, and Outcome is one of the Chapter 11 outcome variables that we examine in Tables 2-3. A selection problem exists if the correlation between the error disturbances of the two equations is not zero, that is, $\text{corr}(\epsilon_i, \eta_i) \neq 0$.

Table 2: Effects of hedge funds on unsecured creditors committee, Probit/OLS

Panel A: Hedge Funds on Creditors Committee

Emerge	Duration	Loss Exclusivity	APR Creditor	Debt Recovery	CEO Turnover	KERP
0.37**	0.17**	0.26	0.38**	0.01	0.16	0.32**
[0.16]	[0.08]	[0.17]	[0.19]	[0.04]	[0.15]	[0.15]

Panel B: Hedge Funds on Equity Committee

Emerge	Duration	Dist Equity	Debt Recovery	Stock Return	CEO Turnover	KERP
0.39	0.16	1.25***	0.18***	0.16***	0.68**	−0.18
[0.30]	[0.15]	[0.28]	[0.07]	[0.04]	[0.27]	[0.29]

Panel C: Hedge Funds Loan-To-Own

Duration	Loss Exclusivity	ARP Creditor	Dist Equity	Debt Recovery	CEO Turnover	KERP
0.05	−0.08	0.67***	0.33**	0.06	−0.14	0.30*
[0.09]	[0.18]	[0.18]	[0.17]	[0.04]	[0.17]	[0.16]

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

This table shows the effect of hedge-fund presence on the unsecured creditors committee, the equity committee, and hedge funds adopting a loan-to-own strategy on Chapter 11 outcome. This includes emergence (Emerge), the logarithm of the number of months in bankruptcy (Duration), the debtor's loss of exclusive rights to file a plan of reorganization after 180 days in bankruptcy (LossExclusivity), APR deviation for secured creditors (APRCreditor), average recovery rate of all corporate debt at plan confirmation (DebtRecovery), CEO turnover during Chapter 11 reorganization (CEOTurnover), adoptions of key employee retention plan (KERP), equity holders receiving positive payoffs (DistEquity), and standardized equity abnormal monthly returns from two days before filing to plan confirmation (StkRet). This table presents results from a simple probit (when the outcome variable is binary) or an OLS (when the outcome variable is continuous) regression model.

Table 3: Effects of hedge funds on unsecured creditors committee, binary outcome with a binary endogenous explanatory variable model/treatment regression

Panel A: Hedge Funds on Creditors Committee

Emerge	Duration	Loss Exclusivity	APR Creditor	Debt Recovery	CEO Turnover	KERP
0.779	0.365	1.884***	0.743	0.500***	1.306***	−0.085
[1.056]	[0.490]	[0.109]	[1.148]	[0.141]	[0.379]	[1.036]

Panel B: Hedge Funds on Equity Committee

Emerge	Duration	Dist Equity	Debt Recovery	Stock Return	CEO Turnover	KERP
1.21*	−0.55*	−0.33	0.19	0.14**	2.33***	0.280
[0.66]	[0.32]	[0.47]	[0.22]	[0.07]	[0.17]	[0.96]

Panel C: Hedge Funds Loan-To-Own

Duration	Loss Exclusivity	ARP Creditor	Dist Equity	Debt Recovery	CEO Turnover	KERP
0.33	1.60***	−0.18	1.11	0.71***	1.09***	−0.18
[0.57]	[0.29]	[0.77]	[0.74]	[0.07]	[0.39]	[0.78]

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

This table is like the previous table, but presents results from a binary outcome model with a binary endogenous explanatory variable (when the outcome variable is binary) or a treatment regression model (when the outcome variable is continuous).

Distinguishing selection from influence

For identification, we need instrumental variables that effectively predict hedge-fund participation but do not affect outcome variables other than through hedge funds. The first variable is the lagged three-month return on an index of distress-investing hedge funds using data from CISDM. The second variable is the residual from regressing the raw lagged three-month return on the S&P 500 index on the return of index of distress-investing. Both variables are valid instruments because they capture the capital supply conditions of hedge-fund distress investment but are also unlikely to directly impact reorganization outcomes of individual cases due to both the exogeneity of market-wide returns to an individual firm and the lack of autocorrelation in returns. Tables 2

and 3 present key results from the simple probit or OLS regression models without and without instrumentation for HFPart, respectively. They examine the effect of hedge-fund participation on creditors committee, equity committee and loan-to-own strategies on Chapter 11 outcomes.

Our results suggest that hedge-fund presence on the unsecured creditors committee is positively associated with all seven outcome variables, and the effects are significant (at the 5% level) for emergence from bankruptcy, duration in bankruptcy, absolute priority rule (APR) deviations for the secured creditors, and the adoption of a KERP. Once the selection effect is taken into account, the coefficient on Hedge Funds on Creditors Committee becomes significant in the outcome equations for the debtor's loss of exclusive rights to file a plan, debt recovery, and CEO turnover, but loses significance in other outcome equations. Our results from both the un-instrumented and instrumented regressions indicate an interesting combination of investment selection abilities possessed by hedge-fund creditors, as well as the activist roles they play. As skilled investors, hedge funds invest in the unsecured debt of distressed firms that are more likely to offer desirable outcomes for that class of claim holders (including emergence, more frequent APR deviations for secured creditors in favor of unsecured creditors, and retention of key employees). Conversely, the debtor's loss of exclusive rights to file a reorganization plan after 180 days and higher CEO turnover rates also appear to be caused by hedge-fund actions.

The effects of hedge-fund presence on the equity committee share similarities to, as well as exhibit differences from, those related to their presence on unsecured creditors committee. Similar to their creditor counterparts, hedge-fund equity holders are just as vigilant in pushing out failed CEOs. The instrumental variable approach allows us to conclude that hedge funds do not serve on the equity committees randomly. In fact, they target companies with more entrenched management. Hedge-fund presence on the equity committee is associated with a large increase in the probability of a positive distribution to existing equity holders, controlling for firm and case characteristics. This effect is rendered insignificant when the instrumental variable approach is used. Together, these results offer strong evidence in support of hedge funds' ability to pick stocks of distressed firms with better prospects for existing shareholders, but offer less evidence for hedge funds' activist role in making the distribution happen.

The results of hedge funds' loan-to-own strategies appear to be a natural blend of their roles on creditors and on equity committees, consistent with the hedge funds' dual roles first as creditors and then as new shareholders. Hedge funds' loan-to-own strategies are pro-KERP, and are associated with greater distributions to both unsecured creditors and shareholders. The effects are significant on the debtor's loss of exclusivity, debt recovery, and CEO turnover in the instrumented model. All of these relationships indicate that the loan-to-own players act like unsecured creditors in exerting their influence over management. At the same time, they value continuity by retaining companies' key employees given that they have a relatively long investment horizon in firms that emerge from Chapter 11.

Our results thus far suggest that hedge funds are effective in achieving their desired outcomes for the claims they invest in. Most notably, our instrumented results on higher debt recovery and stock returns with hedge-fund activism are more supportive of efficiency gains rather than value extraction

by hedge funds from other claims. Such value creation may come from overcoming secured creditors' liquidation bias (i.e. a higher probability of emergence), confronting underperforming managers (i.e. a higher probability of loss of exclusivity and a higher CEO turnover rate), retaining key personnel (i.e. more frequent adoptions of KERP), and relaxing financial constraints (i.e. the loan-to-own strategy).

Markets recognize hedge fund activism effectiveness

To provide further evidence in support of efficiency gains brought by hedge funds as opposed to value extractions from other claims, we adopt an event study that relates changes in stock prices around the bankruptcy filing to hedge-fund involvement on the debt side that is observable at that time. In our sample of 277 cases, hedge funds are listed among the largest unsecured creditors on the bankruptcy petition forms in 75 cases. Figure 1 plots the cumulative abnormal returns (CARs) of the group with hedge funds as creditors and the group without hedge funds for the [−10, +10] window, where day 0 is the date of the Chapter 11 filing. We show that after the Chapter 11 filing, the group with hedge-fund presence experiences price increases, while the group without any hedge-fund presence continues to experience price declines.

Figure 1: Event study around Chapter 11 filing
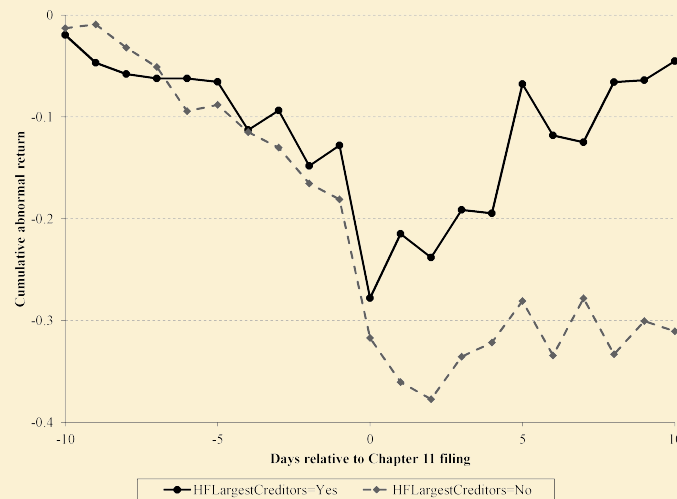
This figure shows the cumulative abnormal returns (CARs, adjusted by the CRSP equal-weighted return) from the 10 days before to the 10 days after a Chapter 11 filing. The solid line represents CARs for 75 cases with at least one hedge fund listed as the largest unsecured creditor. The dashed line represents CARs for 202 cases without any hedge fund listed as the largest unsecured creditor.

Conclusion

To conclude, our study finds that hedge funds' choice in distressed targets and positions in the capital structure reflect both their firm-picking skills and their desire to have a larger impact on the restructuring process. Hedge funds are largely effective in achieving favorable outcomes for the claims that they choose to invest in. Their success in helping distressed firms reorganize does not come at the expense of other claimants, but rather from creating value for the firm as a whole. This study adds to our understanding of the major forces underlying the patterns of, and changes in, the Chapter 11 process in the United States over the past decade. Additionally, it contributes to the growing research on investor activism. By analyzing hedge-fund holdings across the capital structures of firms in Chapter 11 restructuring, our work also stimulates new theoretical research on bankruptcy that allows for complex and dynamic interactions among the variety of relevant stakeholders.
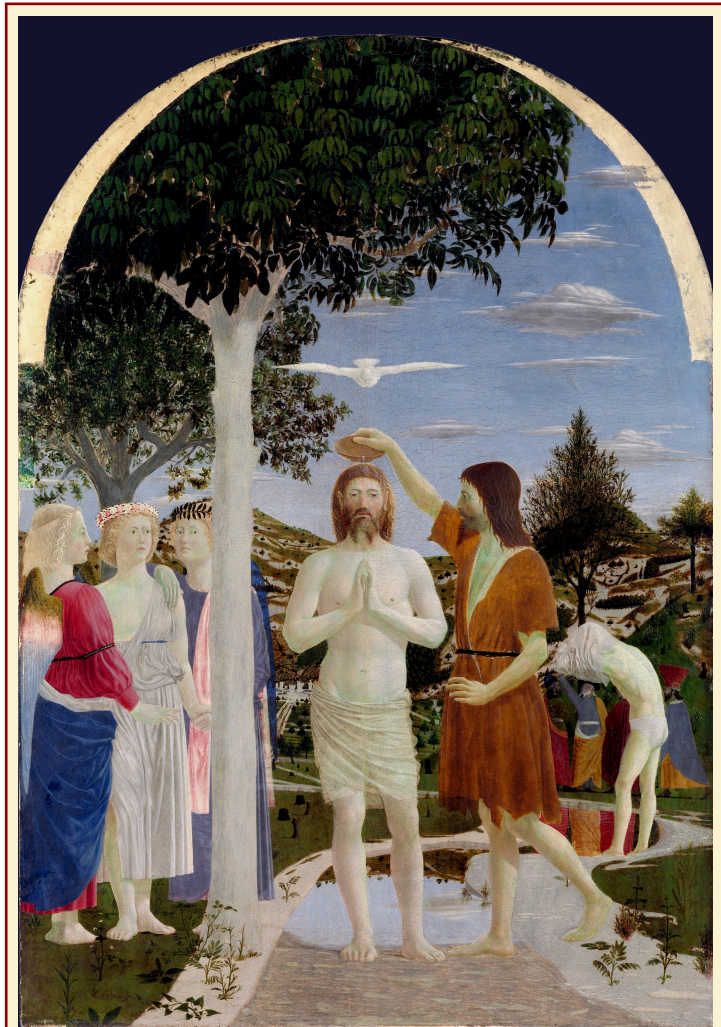
Piero della Francesca (1415-1492): The Baptism of Christ. Italy, 1450. A mind-expanding use of space and geometry, centuries ahead of its time. What do you notice when you first look at the painting? Jesus? The onlookers? The white and blue colors? Bet it is not the single biggest thing in the painting—the foliage!

Oldrich Alfons Vasicek

General equilibrium with heterogeneous participants and discrete consumption times

Vasicek (JFE 2013) presents a computable general equilibrium model that relates interest rates to the underlying economic variables such as the risks and returns of real production, the risk tolerances and time preferences of the investors, and the distribution of wealth in the economy. The model provides a means of quantitative analysis of how economic conditions and scenarios affect interest rates.

The continuous-time economy includes risky production subject to uncertain technological changes. Consumption takes place at a finite number of discrete times. Each investor maximizes the expected utility from lifetime consumption. The participants have constant relative risk aversion, with different degrees of risk aversion and different time preference functions.

An equilibrium with heterogeneity

For a meaningful economic analysis, it is essential that a general equilibrium model allows heterogeneous participants. If all participants have identical preferences, then they all hold the same portfolio. Because there is no borrowing and lending in the aggregate, there is no net holding of debt securities by any participant, and no investor is exposed to interest rate risk. Moreover, if the utility functions are the same, it does not allow for study of how interest rates depend on differences in investors' preferences.

The main difficulty in developing a general equilibrium model of production economies with heterogeneous participants had been the need to carry the individual wealth levels as state variables, because the equilibrium depends on the distribution of wealth across the participants. This had precluded an analysis of equilibrium in a production economy with any meaningful number of participants; most explicit results for production economies had previously been limited to models with one or two participants.

The paper shows that the individual wealth levels can be represented as functions of a single process, which is jointly Markov with the technology state variable. This allows construction of

equilibrium models with just two state variables, regardless of the number of participants in the economy.

Consider a continuous time economy with n participants endowed with initial wealth $W_k(0)$, $k = 1, 2, \ldots, n$. It is assumed that investors can issue and buy any derivatives of any of the assets and securities in the economy. The investors can lend and borrow among themselves, either at a floating short rate or by issuing and buying term bonds. The resultant market is complete. It is further assumed that there are no transaction costs and no taxes or other forms of redistribution of social wealth. The investment wealth and asset values are measured in terms of a medium of exchange that cannot be stored unless invested in the production process.

The economy contains a production process whose rate of return dA/A on investment is

$$\frac{dA}{A} = \mu \cdot dt + \sigma \cdot dy,$$

where y(t) is a Wiener process. The process A(t) represents a constant return-to-scale production opportunity. The amount of investment in production is determined endogenously.

The parameters of the production process can themselves be stochastic, reflecting the fact that production technology evolves in an unpredictable manner. It is assumed that their behavior is driven by a Markov state variable $X(t)$, $\mu = \mu(X(t), t)$, $\sigma = \sigma(X(t), t)$. The state variable can be interpreted as representing the state of the production technology. The process X(t), which can be a vector, may be correlated with the production process A(t).

The consumption is restricted to a set of specific discrete dates $t_1 < t_2 < \ldots < t_m = T$. The economy exists in continuous time, and between the consumption dates the participants are continuously trading and production is continuous. Each investor maximizes the expected utility from lifetime consumption,

$$\max E \sum_{i=1}^{m} p_{i,k} \cdot U_k(C_{i,k}),$$

where $C_{i,k}$ is the consumption at time t_i and U_k is a utility function given by

$$U_k(C) = \frac{C^{(\gamma_k - 1)/\gamma_k}}{\gamma_k - 1}$$

if $0 < \gamma_k, \gamma_k \neq 1$ and log(C) when $\gamma_k = 1$.

An economy cannot be in equilibrium if arbitrage opportunities exist. A necessary and sufficient condition for absence of arbitrage is that there exist a process Y(t), called the state price density process, such that the price P of any asset in the economy satisfies the equation

$$P(t) = E_t \left[P(s) \cdot \frac{Y(s)}{Y(t)} \right].$$

Equilibrium is fully described by specification of the process $Y(t)$, which determines the pricing of all assets in the economy, such as bonds and derivative contracts. Bond prices in turn determine the term structure of interest rates. The state price density process also determines each participant's optimum investment strategy by means of the formula for $W_k(t)$ below. Solving for the equilibrium means solving for the process $Y(t)$.

A simple solution and algorithm

The paper shows that the optimal consumption of the k-th investor is a function of his own preference parameters and the state price density process only, given by

$$c_{i,k} = v_k \cdot p_{i,k}^{\gamma_k} \cdot Y^{-\gamma_k}(t_i)$$

for $i = 1, 2, ..., m$, $k = 1, 2, ..., n$, where v_k is a constant determined by the initial wealth $W_k(0)$. The investor's wealth $W_k(t)$ at time t under an optimal investment and consumption strategy is

$$W_k(t) = v_k \cdot \frac{1}{Y(t)} \cdot E_t \left[\sum_{i:t_i > t} p_{i,k}^{\gamma_k} Y^{-\gamma_k+1}(t_i) \right].$$

In equilibrium, the total wealth $W(t) = \sum W_k(t)$ must be invested in the production process (the market portfolio). Any lending and borrowing, including lending and borrowing implicit in issuing and buying contingent claims, is among the participants in the economy, and its sum must be zero. This requirement produces equations for the state price density process $Y(t)$.

These equations have a unique solution for $Y(t_1), Y(t_2), \ldots, Y(t_m)$ in the form

$$Y(t_i) = F_i[Y(t_{i-1}), A(t_{i-1}), X(t_{i-1}), A(t_i), X(t_i)],$$

where F_i for $i = 1, 2, \ldots, m$ are functions determined by the algorithm. These variables in turn specify the state price density process $Y(t)$ in continuous time. The algorithm requires no more complicated mathematical tools than finding the root of a monotone function. This represents the exact solution to the equilibrium economy, provided the initial wealth distribution is as specified.

To determine the values of the constants $v_1, v_2, \ldots, v_n$, our paper utilizes the fact that any choice of the constants is consistent with a unique equilibrium described by the process $Y(t)$, except that the initial wealth levels $W_k'(0)$ obtained from the equation for the optimal strategy for $t = 0$ do not agree with the given initial values $W_k(0)$. Repeatedly replacing v_k by $v_k \cdot W_k(0)/W_k'(0)$ and recalculating $Y(t)$ converges to the required equilibrium. The proof of convergence is given in the paper.

While our paper concentrates on the case that the participants have iso-elastic utility functions, the approach can be extended to more general class of utility functions.

Hieronymus Bosch (1450–1516): The Conjurer. Flemish, 1502. "No one is so much a fool as a willful fool." This is the proverb depicted in this Bosch painting. It shows how people are fooled by their lack of alertness and insight. If it was today, could Bosch have painted the subprime crisis from the bankers' point of view?

Christian C. Opp, Marcus M. Opp, and Milton Harris

Rating agencies in the face of regulation

> "The story of the credit rating agencies is a story of colossal failure."
> Henry Waxman (D-CA), Chairman of the House Oversight and Government Reform Committee.

Massive downgrading and defaults during the 2008/2009 financial crisis have led politicians, regulators, and the popular press to conclude that the rating agencies' business model is fundamentally flawed. The popular argument goes as follows. Issuers of securities naturally prefer higher ratings for their issues, because these reduce their cost of capital. Because the issuer pays the rating agency to provide a rating, rating agencies can capture some or all of the benefit to the issuer of providing high ratings. This results in "huge conflicts of interest" (Krugman, NYT 2010) between rating agencies and investors, because rating agencies have an incentive to inflate their ratings relative to the information available.

Paying for ratings does not always work

Recent academic studies provide a more nuanced perspective. For example, although rating standards in the residential mortgage-backed securities (MBS) market declined in the years leading up to the 2008/2009 crisis (Ashcraft-Goldsmith-Pinkham-Vickery, WP 2010), they stayed conservative for corporate bonds. Similarly, exotic, structured securities receive a much higher percentage of AAA ratings (e.g., 60% for collateralized debt obligations or CDOs) than do corporate bonds (1%, see Fitch, Brochure 2007: "Inside the ratings," Information brochure, Fitch). These facts are difficult to explain based purely on conflicts of interest inherent in the issuer-pays model. Moreover, the simple conflict-of-interest story ignores the importance of reputation for credit rating agencies.

Mechanical regulator rating use is to blame

But if it is not the conflict of interest between investors and credit rating agencies that caused the apparent inflation in ratings during the financial crisis, what does explain this event? We argue the culprit is the mechanical use of ratings in government regulation of financial institutions. The cost of regulatory compliance for these institutions, e.g., capital and reserve requirements, is reduced to the extent that they invest in highly rated securities instead of lower rated securities. Some or all

of the regulatory cost reduction can be captured by the rating agency in the form of higher fees for a rating. This outsourcing of regulatory risk assessment to credit rating agencies provides them with a source of revenue that is unrelated to the informativeness of the rating which can produce incentives for rating inflation that are not abated by the agencies' reputation concerns. Even if rating agencies cannot fool investors into believing that higher rated securities are less risky, they can still capture (some of) the regulatory cost reduction for higher ratings by issuing more high ratings. (Incorporating the regulatory use of ratings into the analysis is also appealing because there is extensive empirical evidence that regulatory implications of ratings are a first-order concern for marginal investors; that is, ratings affect market prices through the channel of regulation, independent of the information they provide about the riskiness of securities (Kisgen-Strahan (RFS 2010) and Ashcraft-Goldsmith-Pinkham-Hull-Vickery (AER 2011)).)

Our model

We incorporate our idea in a simple model that includes a large number of potential issuers (firms), a rating agency, and a large number of regulated investors who compete to purchase the issued securities. Issuers of securities within a given class (e.g., corporate bonds or mortgage backed securities) are of heterogeneous quality. In particular, there are two types of issuers, "good" issuers with positive NPV projects and "bad" issuers with negative NPV projects. The average NPV of all issuers is assumed to be negative. Each issuer knows its own type, but this information is not available to the credit rating agency or investors. The credit rating agency tries to identify the types (good or bad) of issuers by collecting costly information. The information technology generates a noisy binary signal (A or B), whose accuracy is larger the more the rating agency spends on information acquisition. Given its information, the rating agency assigns one of two ratings, A or B, to an issuer that requests a rating. The rating may be the same as the signal or not, e.g., the rating agency can report an A rating even if its signal is B.

As is the case in reality, issuers may request a rating for free, but, after seeing the rating, must choose whether to pay the rating agency to publish it. They will do so if and only if the rating allows the firm to sell the issue to investors at a price that covers the firm's investment cost plus the rating agency's fee. Regulated investors purchase issues at this price if and only if they expect at least to break even on the purchase, including any savings in regulatory compliance if the issue receives an A rating (investors will not purchase B-rated securities, because these will have negative NPV and offer no savings in regulatory compliance). We refer to the investors' savings in regulatory compliance as the regulatory advantage, denoted y, of an A rating.

The rating agency chooses its fee, how much to spend on generating information, and how to assign ratings given the resulting information signal to maximize its expected revenue from selling ratings net of its cost of generating the signal, given the behavior of firms and investors. Having a monopoly on issuing ratings, the rating agency sets its fee to make investors just willing to purchase A-rated securities and firms just willing to purchase A ratings. If ratings are sufficiently informative, the average project NPV of the firms with A ratings becomes positive (i.e., a high enough fraction

of A-rated firms is good). In this case, the rating agency captures the regulatory advantage from investors and some of the NPV of the A-rated firms' projects.

Rating inflation can make sense under regulation

It is easy to see that the rating agency, if it invests in producing information, will report its signal truthfully, because otherwise there is no point in producing the information. The rating agency will, therefore, pursue one of two strategies: **rating inflation**, i.e., produce no information at zero cost and rate all issues as A; or **full disclosure**, i.e., produce some information, report it truthfully, and sell ratings on all securities that receive A signals. The advantage of rating inflation is that it maximizes the volume of ratings sold and has no information cost. The disadvantage is that the NPV of the resulting A-rated projects will be the average project NPV in the population, which is negative, reducing the agency's fee below the regulatory advantage y of A-rated securities. Thus, rating inflation will be optimal only when the regulatory advantage is large relative to the cost of producing information and the average NPV of the firms in the population. We denote the threshold for the regulatory advantage above which the rating agency pursues rating inflation by $\bar{y}$. It is important to note that when $y \leq \bar{y}$, there is no attempt by the rating agency to inflate ratings, despite the fact that the issuer pays for its own rating. Rating agency profits for the two strategies are plotted in Figure 1.

Figure 1: Full disclosure vs. rating inflation

The graph plots profits under full disclosure $\pi_{FD}(y)$ and rating inflation $\pi_{RI}(y)$ as a function of the regulatory advantage y. The rating inflation threshold, $\bar{y}$, for this example is 0.24. Equilibrium profits $\pi^*(y)$ for $y < 0.24$ are attained by full disclosure. At $y = \bar{y} = 0.24$, profits from full disclosure and rating inflation are equal. Rating inflation obtains for $y > 0.24$.

More complex securities are more inflated

The threshold $\bar{y}$ is determined by the cost of producing information and the characteristics of the issuers, such as the fraction of good types, denoted π_g, and the profitability of their projects. If the regulatory subsidy for a high rating is close to the threshold, small changes in these parameters can have a huge impact on the informativeness of ratings. In particular, we show that the threshold $\bar{y}$ decreases with increases in the cost of information, the fraction of good issuers in the population, and the profitability of a good project. (Figure 2 illustrates the first two results.) Thus a small increase in any of these characteristics, as well as a small increase in the regulatory subsidy itself, can result in ratings that change from informative to totally useless. In particular, it seems plausible that newer, more complex securities are more costly to rate than traditional securities (such as corporate bonds) for which rating agencies have considerable experience. Our result then implies that ratings of newer, more complex securities may be inflated, while those of traditional securities are not.

Figure 2: Regulatory advantage and equilibrium information quality
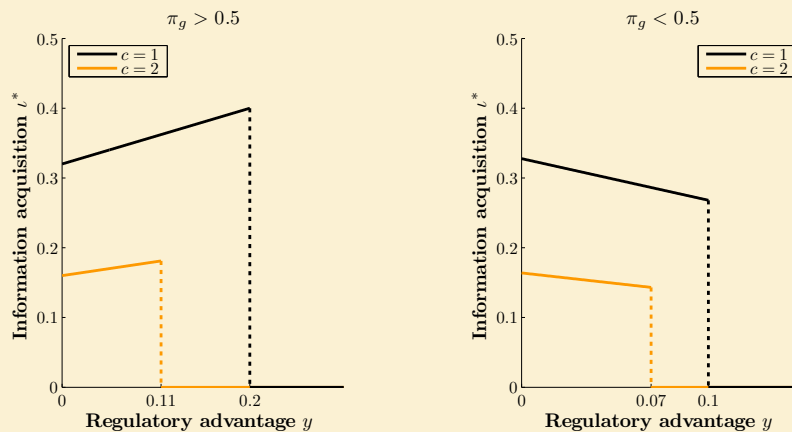
Both panels illustrate the comparative statics of regulatory advantage $\bar{y}$ and ι^* with respect to marginal information costs c, where $C'(\iota) = c \cdot \iota$ and ι is the quality of the rating agency's signal. The inflation threshold $\bar{y}$ falls when c increases from 1 to 2 (from 0.2 to 0.11 in the left panel and from 0.1 to 0.07 in the right panel). Equilibrium information quality also falls, for any given regulatory advantage when the marginal cost of information increases. The left (right) panel shows the effect of changes in the regulatory advantage on information quality when the proportion of good types is more (less) than 1/2, respectively. Increases in the regulatory advantage increase information quality when there are more good types than bad and vice versa when there are more bad types than good.

When are ratings informative?

When the regulatory subsidy is not so large as to result in rating inflation, the amount of information produced by the rating agency still varies with the size of the regulatory subsidy as well as with the cost of information and the issuers' characteristics. In particular, an increase in the regulatory subsidy increases the informativeness of ratings, all else equal, if there are more good

issuers than bad issuers ($\pi_g > 1/2$) but reduces informativeness if there are more bad issuers than good ($\pi_g < 1/2$). This is because, when the regulatory subsidy is increased, the rating agency's incentive to rate issues highly increases. Given that the subsidy is small enough so that the rating agency reports truthfully, it will rate more issues highly by rating them more accurately if and only if there are more good issues than bad ones. Not surprisingly, a higher marginal cost of information production decreases the informativeness of ratings. These results are illustrated in Figure 2. The first is seen by comparing the left and right panels. In the left panel of Figure 2, information acquisition increases as a function of y, because the fraction of good types, π_g, is $0.7 > 1/2$. In contrast, the right panel plots a case in which the fraction of good types is $0.2 < 1/2$ so that information acquisition decreases. The second result is shown in both panels upon comparing low ($c = 1$) and high ($c = 2$) marginal information costs.

To turn these comparative statics results into testable predictions, we must first relate the model parameters to their empirical counterparts. First note, we should interpret our signals A vs. B relative to publicly available information, e.g., conditional on the size/leverage of the firm and the security class. (This is consistent with the behavior of actual rating agencies which generally provide relative assessments within particular security classes, rather than across security classes. For example, for some firms, the distinction between A and B in our model would refer to the difference between investment-grade and junk status, while, for others, it would represent the difference between Aa and A.) In particular, following the results by Kisgen-Strahan (RFS 2010) and Ellul-Jotikasthira-Lundblad (JFE 2011), the regulatory advantage y is especially large around the investment grade / junk threshold and at the AAA vs. AA threshold. (Kisgen-Strahan (RFS 2010) estimate that the reduction in the debt cost of capital is 54 bps around the investment grade cutoff vs. an average reduction of 39 bps.) This leads to strong incentives to inflate around these thresholds, implying a large drop in rating informativeness.

Secondly, while the previous source of variation in y results from differential regulatory importance across rating grades, differential importance of regulation to the marginal investor across security classes provides cross-sectional variation in y. To the extent that the marginal investor's regulatory constraint binds in one security class (say CMBS, in which the marginal investor is an insurance company), but does not bind in another security class (e.g., MUNI, in which the marginal investor is a retail investor), one would expect cross-sectional differences in the incentives to inflate. Similarly, one could exploit cross-sectional variation in the "tightness" of regulatory constraints across countries. To our knowledge, neither of these avenues has been explored in the empirical literature so far.

Third, time-series changes in regulation provide quasi-natural experiments. Here, one can distinguish between changes in regulation of institutional investors, as exploited in the CMBS sample of Stanton-Wallace (WP 2010), or changes in the regulatory status of a rating agency, such as in Kisgen-Strahan (RFS 2010). In the former case, our analysis predicts the rating inflation in the CMBS market documented in Stanton-Wallace (WP 2010). In the latter case, Kisgen and Strahan investigate empirically the results of the SEC's accreditation of Dominion Bond Rating Services as an NRSRO. This accreditation allowed Dominion's ratings to be used for regulatory purposes,

implying that only post accreditation, a high rating by Dominion offered a regulatory advantage, i.e., y > 0. Consequently, our model predicts a shift in the distribution of Dominion's assigned ratings towards better ratings, especially around the relevant cutoffs, post SEC accreditation.

Finally, our model also has implications for the planned overhaul of financial regulation. In contrast to the supranational Basel III guidelines, the recently proposed Dodd-Frank Act aims to eliminate all regulation based on ratings in the U.S. If this fundamental regulatory change is implemented, we would expect a reduction of the regulatory advantage of higher ratings. As a result, our model would predict a systematic downward shift in the distribution of ratings of the current NRSROs, especially around the two identified thresholds. Whether abandoning rating-contingent regulation is preferable from society's perspective depends on the alternatives to rating-contingent regulation, in particular how Dodd-Frank's mandate to use "all publicly available information" is implemented. (For example, the national insurance regulator NAIC started to use risk assessments by market participants (Pimco and Blackrock) for capital regulation of insurance companies (see Becker-Opp (WP 2013)).) We leave this question for future work (Harris-Opp-Opp (WP 2013)).

Leonardo de Vinci (1452–1519): Lady with an Ermine. Italy, 1490. This portrait was commissioned by the Duke of Milan, immortalizing his beautiful and scholarly mistress Cecilia, who he never married although they had a son. In fact, he said that he "no longer wished to touch [Cecilia] or have her nearby, because she was so fat, ever since giving birth." How things have changed for age-old beauties!

Two wrongs make a right: when consistency trumps accuracy in forecasting.

It turns out that being wrong may be the right career move for financial analysts who want to move stock prices and markets while also moving up the professional ranks.

This curious fact emerged in our recent research of analyst forecasting. Our study found that forecast consistency, rather than accuracy, provided more useful information to a large investor segment—"savvy" investors, as distinct from their less sophisticated, "retail" counterparts. As a result, these analysts could move stock prices and influence the stock market more strongly than analysts who provided more accurate, but inconsistent, forecasts.

Specifically, we found that analysts who strategically introduced a downward bias in their forecasts ("lowballed"), enjoyed higher market credibility by demonstrating a lower standard deviation of forecast error. These analysts often curried favor with management, because managers frequently could outperform lowball forecasts. In return, these managers granted analysts greater access to company information. Consequently, these analysts enjoyed better career advancement and professional recognition than their peers who were inconsistent, if approximately more accurate, in their own forecasts.

In other words, to err is human, but to err with reliable frequency may make you an "All Star" analyst.

For example, consider two forecasts. Analyst A delivers forecasts that are consistently three cents below realized earnings, while Analyst B provides forecasts that are two cents above realized earnings half the time and two cents below realized earnings the other half of the time. Investors should prefer Analyst A's forecasts. Why? Despite their stated lower accuracy compared with the forecasts of Analyst B, Analyst A's forecasts prove more useful as they are a predictable transformation of realized earnings.

In fact, when examining 12 years of data, we found that this "consistency effect" on informativeness is some two to four times greater than the effect of accuracy. (Earnings and analyst forecast data are from the 1994–2006 I/B/E/S Detail History files, specifically quarterly forecasts from analysts with eight or more quarters of experience.)

By shifting the focus of forecast informativeness from "accuracy" to "consistency," our research shows that the volatility of earnings forecast errors can prove more important than their magnitude. This fact has implications for investors, analysts and regulators alike.

For example, importantly, we found that the consistency affect worked in the presence of institutional investors, who functioned as our proxy for sophisticated investors capable of discerning systemic bias and extracting useful information from the biases. Less sophisticated investors, by contrast, were less likely to decipher the bias and so tended to prefer forecast accuracy, even when that accuracy was inconsistently delivered. Indeed, when investors failed to recognize systemic bias, they penalized analysts for issuing (consistently) inaccurate forecasts.

These results also have implications for analysts' careers. Consistent with our expectations, more consistent analysts are less likely to be demoted to less prestigious brokerage houses and are more likely to become All Stars. We also found that analysts who lowball are more consistent but less accurate. These effects are particularly strong for analysts covering firms with more institutional investors.

Our results also offer useful insights for evaluating legislature such as the Global Settlement of 2003 (that requires research analyst's historical ratings to be disclosed) or the Regulation Fair Disclosure (Reg FD) of 2000 (that requires all public companies to disclose relevant information to all investors simultaneously). Overall, regulation has curtailed selective disclosure, at least to some extent, and in turn decreased analyst lowballing activity, which has resulted in less consistent forecasts. In other words, removing systematic bias, as regulators have done, levels the playing field but reduces the efficiency of price formation.

Empirical design and main results

To test our basic intuition, we first needed a measure of forecast informativeness. Beta is the coefficient obtained by regressing a three-day abnormal stock returns around the forecast revision date on forecast revisions over all quarters for which analyst i covered firm j. We then regressed this measure of forecast informativeness (Beta) on our measures of consistency (Cons) and accuracy (Accu), controlling for different relevant variables. We measured our variables for each firm-analyst over the entire sample period. Cons is a rank measure based on the standard deviation of the forecast errors over all quarters for which analyst i covered firm j, while Accu is a rank measure based on the absolute value of analyst forecast error. Cons and Accu are only moderately correlated (approximately 0.30). Specifically, we estimate the following cross-sectional model for each analyst i and firm j:

$$\text{Beta}_{i,j} = \alpha_0 + \alpha_1 \cdot \text{Cons}_{i,j} + \alpha_2 \cdot \text{Accu}_{i,j} + \alpha_k \cdot X_{i,j}^k + e_{i,j}.$$

> **Table 1: The association of informativeness (Beta) with consistency and accuracy**
>
> | Dependent Variable: Informativeness | | |
> | --- | --- | --- |
> | | Coefficient | (StdErr) |
> | Cons(istency) | −19.69*** | (3.70) |
> | Accu(racy) | 8.71*** | (0.78) |
> | Others : See description | | |
> | $N = 38,096$, $R^2 = 1.93\%$ | | |
>
> Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.
>
> The dependent variable, beta, measures the informativeness of analysts. It is the coefficient obtained by regressing a three-day abnormal stock returns around the forecast revision date on forecast revisions over all quarters for which analyst i covered firm j. An analyst who does not change/update with stock-market changes has a low beta. Cons(istency) is a rank measuring the variability of forecast errors. Accu(racy) is a rank measuring the difference between actual and forecast earnings. The regression includes other variables: Intercept (significant at 1%), Horizon (at 5%), Boldness (at 5%), Brokersize, Experience, Breadth, and Cover.

We report the results of this analysis in Table 1. As predicted the coefficient associated with Cons is more significant than the coefficient associated with Accu. To examine the effect of investor sophistication, we split our overall sample into two subsamples based on the percentage of institutional investor ownership and we reestimate Model (1) separately for each subsample. Untabulated results indicate that the effect of consistency is more significant in the subsample in which sophisticated investors are more present.

To investigate the effect of consistency and accuracy on analyst's career, we estimate the following models:

$$\text{Demo}_{i,t} = \gamma_0 + \gamma_1 \cdot \text{Cons}_{i,t} + \gamma_2 \cdot \text{Accu}_{i,t} + \gamma_k \cdot X_{i,t}^k,$$

$$\text{AllStar}_{i,t} = \delta_0 + \delta_1 \cdot \text{Cons}_{i,t} + \delta_2 \cdot \text{Accu}_{i,t} + \delta_k \cdot X_{i,t}^k,$$

where Demo is an indicator variable that equals one if analyst i is demoted in the following year (zero otherwise) and AllStar is an indicator variable that equals one if analyst i is on Institutional Investor magazine's All Star list (zero otherwise). Results of this analysis reported in Table 2 indicate that analysts exhibiting a high forecast consistency are less likely to be demoted and more likely to be nominated as an All-Star analyst.

Table 2: The effect of consistency and accuracy on analyst demotions and promotions

	Dependent	
Variable	Demotion	AllStar
Cons(istency)	–0.30***	0.60***
	(0.07)	(0.12)
Accu(racy)	–0.03	0.54***
	(0.10)	(0.13)
Others : See description		
N	15,561	11,985
Pseudo R^2	7.57%	23.18%

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

The left column predicts next-year demotions (transfer to a smaller broker). The right column predicts next-year nomination to the All-Star list. Cons(istency) is a rank measuring the variability of forecast errors. Accu(racy) is a rank measuring the difference between actual and forecast earnings. The regression includes other variables: Intercept (significant at 1%), Horizon (at 5%), Boldness (at 5%), Brokersize, Experience, Breadth, and Cover.

El Greco (1541–1614): Adoration of Shepherds. Spain, 1614. El Greco was known for his unique interpretations of known religious themes with his highly expressive style. This style has been said to have inspired the modern expressionists. Picasso's famous Les Demoiselles d'Avignon show his fascination with El Greco's dramatic use of extended figures and the ability to bring them out from the background.

Renhui Fu, Arthur Kraft, and Huai Zhang

The effect of financial reporting frequency on information asymmetry and the cost of equity

In recent years, there have been calls to increase the frequency of reporting financial statements both in the U.S. and in other countries. This raises questions as to what are the effects of increased reporting frequency. In this study, we examine directly how the frequency of interim reporting affects information asymmetry and the cost of equity.

The predictions from the theoretical literature are unclear. While more frequent disclosures may reduce information asymmetry, they may also provide stronger incentives for sophisticated investors to acquire private information and discourage information production from other sources, resulting in wider information asymmetry among investors. Similarly, while some earlier studies suggest that more disclosures lower the cost of equity by reducing adverse selection and estimation risks, two recent studies suggest that the impact of disclosures on the cost of equity exists only when the disclosures convey information on non-diversifiable risks. Consistent with the different views in theoretical works, empirical evidence is mixed on the relation between disclosures and information asymmetry/cost of equity. Therefore, the impact of financial reporting frequency on information asymmetry/cost of equity remains an empirical issue.

The empirical evidence however is difficult to come by because, after 1970, all firms in the U.S. report on a quarterly basis, making it impossible to observe variations in reporting frequency. To overcome this obstacle, we hand collect reporting frequency data for U.S. firms for the years between 1951 and 1973. During these years, there are substantial variations in the reporting frequency because many firms report more frequently than required by the SEC. (The SEC required annual reporting in 1934, semi-annual reporting in 1955 and quarterly reporting in 1970.) By offering substantial cross-sectional and time-series variation in reporting frequency, our sample period provides an ideal setting to investigate our research question.

Table 1: Effects of reporting frequency

Model/Variable	OLS	Fixed Effects	2SLS
A. Bid-Ask Spread	–0.146***	–0.093***	–0.085***
(IA-Spread$_{i,t}$)	(0.047)	(0.020)	(0.018)
B. Price Impact	–0.382***	–0.225***	–0.216***
(IAPI$_{i,t}$)	(0.094)	(0.049)	(0.047)
C. Ex-post realized returns	–1.482***	–0.854**	–0.885**
(RET$_{i,t}$)	(0.489)	(0.385)	(0.370)
D. Expected CAPM returns	–1.085**	–0.655**	–0.628**
(COE-RET$_{i,t}$)	(0.532)	(0.298)	(0.272)
E. Expected FF3 returns	–1.014**	–0.662**	–0.658**
(COE-FF3$_{i,t}$)	(0.459)	(0.294)	(0.274)
F. Earnings-price ratio model	–0.906**	–0.591***	–0.502***
(COE-EP$_{i,t}$)	(0.438)	(0.153)	(0.147)

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

A typical model looks like

$$y = \alpha + \beta_1 \cdot \text{Freq}_{i,t-1} + \beta_2 \cdot \text{Size}_{i,t-1} + \beta_3 \cdot \log(\text{Turnover}_{i,t-1}) + \beta_4 \cdot \log(\text{Volatility}_{i,t-1}) + \epsilon_{i,t}$$

where y is the variable reported in the first two columns, and the reported coefficients are the reporting frequency, possibly IV-fitted. The reported coefficient is on the first variable, the frequency of reporting.

Frequent reporting reduces the cost of equity

Our results based on the pooled sample are reported in Table 1. In a simple OLS regression of information asymmetry/cost of equity on reporting frequency and other control variables, the coefficient on reporting frequency is negative and significant, suggesting that firms with higher reporting frequency have lower information asymmetry/cost of equity. To alleviate the concern that that some unobservable firm characteristics, such as the firm's riskiness, affect both observed reporting frequency and information asymmetry/cost of equity, we also estimate a firm fixed effects model, a two-stage least squares estimation procedure (2SLS hereafter), and a matched control sample approach. We obtain similar inferences from both the firm fixed effects model and two-stage procedure. Specifically, results from the two-stage procedure suggest that an increase of one in the reporting frequency on average reduces our information asymmetry measure, the price impact, by 0.216% and the cost of equity measure based on the CAPM model by 0.628%.

Table 2: Results based on the matched control sample

Panel A: Voluntary increases in reporting frequency (N = 1,090)

Variables	Information asymmetry			Cost of equity		
	IA_{Spread}	IA_{PI}	COE_{RET}	COE_{CAPM}	COE_{FF3}	COE_{EP}
Treatment×After	−0.162**	−0.431***	−1.613**	−1.217***	−1.308*	−0.963**
	(0.078)	(0.152)	(0.727)	(0.323)	(0.667)	(0.442)

Panel B: Mandatory increases in reporting frequency (N = 1,258)

Variables	Information asymmetry			Cost of equity		
	IA_{Spread}	IA_{PI}	COE_{RET}	COE_{CAPM}	COE_{FF3}	COE_{EP}
Treatment×After	−0.171**	−0.455***	−1.644***	−1.279***	−1.329**	−1.049**
	(0.076)	(0.116)	(0.611)	(0.215)	(0.544)	(0.425)

Panel C: Decrease in reporting frequency (N = 702)

Variables	Information asymmetry			Cost of equity		
	IA_{Spread}	IA_{PI}	COE_{RET}	COE_{CAPM}	COE_{FF3}	COE_{EP}
Treatment×After	0.102	0.381	1.053	1.224	1.255	0.874
	(0.159)	(0.240)	(0.789)	(1.188)	(0.909)	(0.553)

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

The table reports only OLS coefficient estimates (multiplied by 100) and firm-year-clustered standard errors for Treatment×After vs a control sample. The dependent variable is an information asymmetry (cost of equity measure). The treatment are three years of data from firms that change the reporting frequency. Control firms have zero. The sample sizes are 3,050 treatment and control firms, with between 550 and 1,100 effective observation per regression from the 1951–1973 period.

Results from the matched control sample approach (reported in Table 2) show that information asymmetry and the cost of equity decrease significantly for firms that increase their reporting frequency relative to control firms, regardless of whether the increase in reporting frequency is voluntary or mandatory. Specifically, the price impact decreases by 0.431% and 0.455% on average and the cost of equity based on the CAPM model drops by an average of 1.217% and 1.279% for firms with a voluntary increase and for firms with a mandatory increase in reporting frequency, respectively. Most increases amount to doubling the reporting frequency (i.e., from semi-annual

reporting to quarterly reporting). Our results related to decreases in reporting frequency are much weaker, possibly because decreases in reporting frequency are typically temporary and do not reflect a commitment to reduced disclosures. Statistically, more than 90% of firms with a reduction in the reporting frequency revert back to the original level or higher level of reporting frequency over the three years after the reduction.

Conclusion

By showing that higher financial reporting frequency reduces information asymmetry and the cost of equity, our study documents the benefits of providing more frequent financial reporting. In particular, our results related to mandatory changes in reporting frequency suggest that these benefits remain, even when firms are forced to deviate from their chosen reporting frequency. We however cannot conclude that firms should be forced to report more frequently, because our analysis does not address the potential costs of increasing reporting frequency (e.g., out-of-pocket costs, proprietary costs). A more detailed analysis of these costs is a fruitful area for future research.

Rembrandt (1606–1669): The Anatomy Lesson of Dr. Nicolaes Tulp. Flemish, 1632. At the age of 26, Rembrandt was already using strong lights and heavy shadows. Anatomy lessons were social events in those days (much like finance seminars these days). The event took place in a theatre with spectators who bought tickets to see dissection of a body. The body in the painting is that of a criminal convicted of armed robbery and sentenced to death by hanging. A solemn event, indeed, enjoyed by many! Did it reduce the incidence of armed robbery, increase scientific knowledge, or just entertain?

A simple way to estimate bid-ask spreads from daily high and low prices

Our paper derives and tests a new way to estimate bid-ask spreads from high and low prices. The estimator is simple to compute and accurate, allowing it to be used in a variety of research contexts. The idea behind the estimator is simple. As shown by Beckers (JB 1983) and Parkinson (JB 1980), the expected value of the log of the high-low price ratio is proportional to the standard deviation of the true value of the security. However, in the presence of bid-ask spreads, the highest transaction price over a trading day will be a buyer-initiated trade at the ask price and the lowest transaction price over a trading day will be a seller-initiated trade at the bid price. As a result, the expected value of the high-low price ratio is a function of the standard deviation and the bid-ask spread. To disentangle the spread and variance portions of the high-low price range, we calculate the sum of the squared log price ranges over two consecutive days,

$$\beta = \sum_{j=0}^{1} \left[\ln\left(\frac{H_{t+j}^{O}}{L_{t+j}^{O}} \right) \right]^2$$

and the squared log of the two-day price range

$$\gamma = \sum_{j=0}^{1} \left[\ln\left(\frac{H_{t,t+j}^{O}}{L_{t,t+j}^{O}} \right) \right]^2$$

where H_i^{O} is the observed high price on day i and L_i^{O} is the observed low price on day i. The sum of the log price ratios over two days contains twice the daily variance and twice the bid-ask spread. The log price ratio for the two-day period contains twice the daily variance, but only one bid-ask spread. Making use of previous work on high-low price ratios, we can set up two equations to solve for two unknowns: the security's standard deviation and its bid-ask spread. These equations can be solved numerically. Alternatively, if we ignore Jensen's inequality, we obtain a closed-form solution

for the bid-ask spread (S), as follows:

$$S = \frac{2 \cdot (e^{\alpha - 1})}{1 + e^{\alpha}},$$

where

$$\alpha = \frac{\sqrt{2 \cdot \beta} - \sqrt{\beta}}{3 - 2 \cdot \sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2 \cdot \sqrt{2}}}$$

(Simulation results suggest that this simplification has little impact on the performance of the estimator. At the same time, the resulting closed-form solutions lead to a substantial reduction in estimation complexity. It is also important to note that this derivation produces an estimate of the standard deviation, in addition to the bid-ask spread. See the original paper for details.) Using these equations, spread estimates can be obtained for each consecutive two-day period. We find that averaging two-day estimates across a month produces reasonably accurate spread estimates for U.S. common stocks. Averaging across longer periods may reduce sampling error, but assumes that the spread and volatility are constant over the longer window.

This technique for estimating spreads can be used with high and low prices over any time interval, not just trading days. In fact, because variances increase with the length of the interval while spreads do not, the signal-to-noise ratio from intraday periods is greater than that from daily periods. We also note that the use of the high-low ratio does not require that trades be reported in the correct sequence, only that high and low prices are reported in the correct time interval.

Complications in using the estimator in practice

The high-low spread estimator relies on the assumption that the variance over a two-day period is equivalent to the sum of two-consecutive single day variances. One reason this assumption may be violated is that markets for most securities are closed overnight. Hence the high-low ratio over a two-day period includes the overnight variance, while the single-day ratios do not. In our work with CRSP data, we find that a simple adjustment for overnight returns works well. Specifically, if the low on the second day exceeds the close on the first day, we assume the price rose overnight by the difference between the close and the low, and vice-versa, and adjust the day 2 high and low by this estimated price change.

There are other reasons why the two day variance may differ from the sum of two one day variances. Infrequent trading of some securities may cause the observed high-low range to be narrower than the true high-low range. In extreme cases, the high and low price for a day can be equal if a security trades only once or only a handful of times during a day. For some assets, price limits may also result in two day variances that are more than twice as large as one day variances. Our Journal of Finance paper discusses ways to deal with these complications.

It is important to note that the high-low estimator may capture price pressure in addition to the bid-ask spread. Specifically, if the high price results from a large trade that executes above the quoted ask (and vice-versa), this price pressure will be captured by the high-low spread estimator.

Price pressure effects may be particularly important during illiquid periods or periods when quoted depths are very small.

Results

The high-low spread estimator can be used for any market and any time period for which high and low trade prices are available. One important advantage of the estimator is that it can be used to obtain transaction cost estimates during historical periods for which intraday trade and quote data are not available. To illustrate this potential application, we examine high-low spread estimates during the period from the Great Depression through World War II. Another important use of the estimator is to estimate transaction costs during periods when intraday data are available, but are difficult to use. The increase over time in the use of computerized trading systems has resulted in a substantial increase in the number of quotes and a corresponding increase in the size of quote databases. This data proliferation makes handling quote data difficult and may also lead to significant problems with matching trades to quotes. In this setting, the high-low spread estimator provides a simple estimate of transaction costs that does not require the use of quote data. To illustrate this application, we examine high-low spread estimates during the recent financial crisis.

For both periods, we use daily high and low prices from CRSP to calculate monthly spread estimates for all available exchange-listed common stocks. We categorize stocks into quintiles based on market capitalization, where cutoffs are defined at the beginning of each month based on NYSE breakpoints. Panel A of Figure 1 shows the average monthly high-low spread estimates across all stocks in each quintile.

As shown in Figure 1, transaction costs rose sharply during the Great Depression. Spreads begin to rise in October 1929 and exhibit a sharp spike from mid-1932 through mid-1933. Average spreads rise as high as 30% for the smallest quintile of stocks and to more than 8% for the middle size quintile. This figure illustrates that the market exhibited a sharp decrease in liquidity during the Great Depression that continued through much of World War II. The high-low estimator provides researchers with a simple and accurate means to study transaction costs during historical periods, such as the Great Depression, where intraday data are not available.

Panel B of Figure 1 plots mean transaction costs during the recent financial crisis. For all size quintiles, spreads begin to rise in August of 2007, peaking from about October 2008 through March 2009. Average spreads for the smallest quintile reach as high as 4.0%. However, increased transaction costs are also evident for the largest stocks, with spreads for these stocks reaching over 2.0%. Spreads appear to decrease by late 2009, though there are sharp increases around the "Flash Crash" in May 2010 and again in August 2011. Notably, the high-low spread estimator allows us to study these patterns without making use of any intraday quote data.

Figure 1: Estimated bid-ask spreads
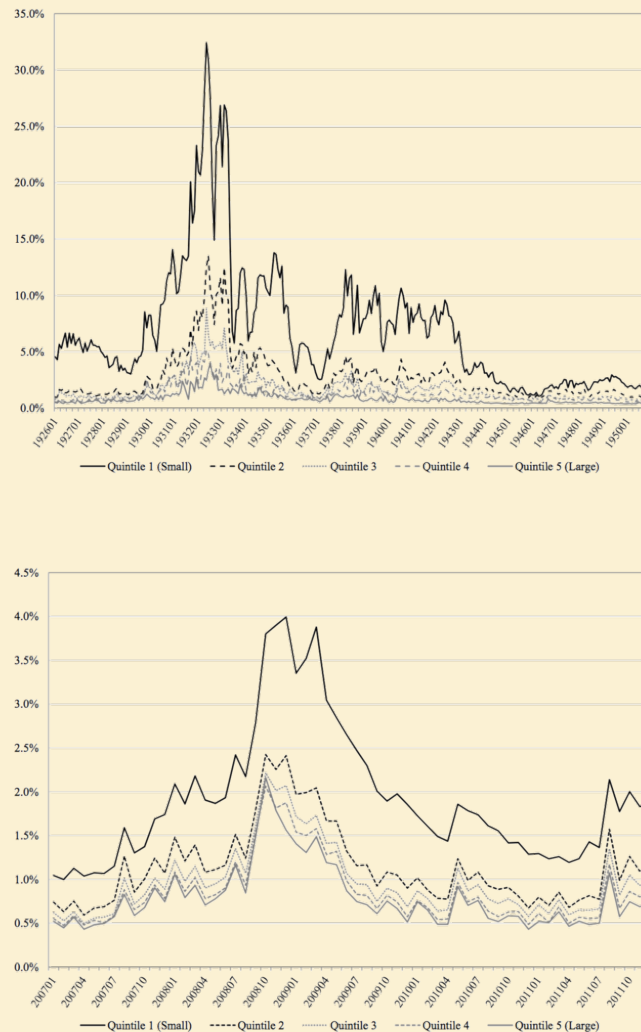
The figure plots average high-low spread estimates across all available common stocks, where stocks are categorized into quintiles based on market capitalization using NYSE breakpoints. Panel A graphs spread estimates through the Great Depression and World War II. Panel B graphs spread estimates through the financial crisis.

As a simple illustration of the estimator's performance, we examine the time-series correlation between market-wide average spread measures based on the high-low estimator and intraday TAQ data. (In our original paper, we provide a detailed analysis of the accuracy and performance of the high-low spread measure relative to alternative transaction cost proxies. We find that the high-low spread estimator generally dominates other low frequency spread estimators at capturing both the cross-section and time-series of individual stock spreads. We also find that the estimator works

best for small, illiquid stocks and during time periods when the minimum tick size is wide.) The spread measure from TAQ is a monthly average of daily time-weighted NBBO quoted spreads. For both measures, we calculate an equal weighted average each month across all available NYSE, Amex, and Nasdaq listed common stocks. Table 1 reports summary statistics and time-series correlations between the market-wide measures for the period from 1993 through 2011 and for various subperiods.

Across the full sample period, the mean (median) high-low spread is 2.18% (2.06%). This compares to a mean (median) TAQ quoted spread of 2.74% (2.54%). This suggests that the high-low estimator slightly underestimates the market-wide quoted spread across the full sample period. The estimator captures the time-series variation in market-wide quoted spreads very well, with a full-period correlation of 0.978. The subperiod results suggest that the high-low estimator underestimates TAQ quoted spreads in the 1990s when then minimum tick size was $0.125 and overestimates quoted spreads during the financial crisis. The overestimation in latter period may reflect, in part, the increased price pressure effects that are captured by the high-low estimator but are not reflected in the quoted spread. As a whole though, the high-low spread estimator provides an accurate and simple way to estimate spreads from daily data.

Table 1: Summary statistics for market-wide spread measures

Period	Number of Months	Time-Series Correlation	TAQ Spread Mean	Median	High-Low Spread Mean	Median
1993–2011	228	0.978	2.74	2.54	2.18	2.06
1993–2000	96	0.972	4.44	4.69	3.09	3.07
2001–2006	72	0.988	1.64	1.14	1.51	1.22
2007–2011	60	0.972	1.33	1.19	1.52	1.39

The table provides summary statistics for market-wide spread estimates based on intraday TAQ data and the high-low spread estimator. Monthly time-weighted quoted spreads based on intraday TAQ data and monthly high-low spreads are estimated for each NYSE, Nasdaq, and Amex listed common stock as described in Corwin-Schultz (JF 2012). Market-wide spreads are then defined each month as an equal weighted average across all available securities.

Raphael (1843–1520): Sistine Madonna. Italian Renaissance, 1502. The hesitant-yet-confident Madonna in this painting is believed to be Margherita, Raphael's mistress, who posed for at least six of Raphael's Madonnas. This painting hangs in Dresden, Germany, since the 18th century. From 1945 to 1955, it was in the Soviet Communist hands before it was returned to the Demokratische Deutsche Republik. They probably couldn't agree whether she was a counter-revolutionary, anyway.

Alex Boulatov and Thomas J. George

Hidden and displayed liquidity in securities markets with informed liquidity providers

Competitive pressure from dark pools has led securities exchanges to offer a variety of ways by which traders can hide orders that provide liquidity. These are the price-contingent orders (limit orders) that aggregate to the market's supply schedule, and against which orders to trade at the market price (market orders) are executed. Hidden orders account for about 30% of transaction volume on some exchanges, and their proliferation raises questions for exchanges and regulators concerning the effect of hidden orders on market quality. In classic microstructure models, adverse selection arises because informed traders can conceal their market orders among the orders of the uninformed. The intuition from these models suggests that explicitly allowing hidden orders will increase adverse selection and the losses suffered by uninformed traders.

We show that the opposite happens when informed traders can choose between providing liquidity by submitting limit orders and demanding liquidity by submitting a market order. In our model, a market with hidden orders imposes smaller losses on the uninformed and has more informative midquotes than a market in which orders that provide liquidity are displayed. This happens because informed traders are drawn into liquidity provision by the incremental profit associated capturing the bid-ask spread when orders are hidden. When orders are displayed, however, some informed traders are deterred from providing liquidity because display expropriates their informational advantage. Competition is more intense when more traders provide liquidity, so the losses to the uninformed are less and prices are more informative when orders are hidden than when they are displayed.

Our model

Our model resembles Kyle (ECTA 1985) and Kyle (RES 1989). It features M risk-neutral strategic informed traders who each observes v_i, a component of the security's payoff $v = v_1 + \ldots + v_M$. There is no competitive market maker, however. Instead, liquidity provision is endogenous. The informed traders choose whether to provide liquidity by placing a price-contingent supply schedule into the order book (a bundle of limit orders), or to demand liquidity by submitting a market order that executes against the book. There are also uninformed liquidity traders who submit net market orders of u. All the random variables are mutually independent and normally distributed.

Each informed trader selects his order to maximize expected profit conditional on the order type he chooses. The number of traders who choose to provide liquidity in equilibrium, J^*, is such that no informed trader can earn greater expected profit by switching his order type. This structure is novel and, we believe, unique to our model. It allows an endogenously determined number of informed liquidity providers and informed liquidity demanders to coexist in the market. In addition, the model can be solved in closed form for an equilibrium with linear strategies that is unique in the linear class.

We compare the equilibria in two market structures. In the first, orders that provide liquidity are hidden from view. In the second, those orders are displayed to market-order traders who observe the contents of the book before choosing their orders. In both market structures, the informed are drawn into liquidity provision to capture rents to providing liquidity—i.e., to earn the bid-ask spread.

Competition is more intense when orders are hidden

When the book is hidden, these rents draw all the informed into trading as liquidity providers, $J^* = M$. It turns out that competition is more intense among the informed when they trade as liquidity providers than when they trade as liquidity demanders. In the former case, they compete on quantity at every price point, whereas in the latter case they compete on only a single quantity across all prices. As a group, the informed would be better off if some could be constrained away from providing liquidity to limit the intensity of competition. But it is not individually rational for any one of them to forgo the rents to liquidity provision. This implies that across all possible allocations of informed traders to order types, the allocation that arises at equilibrium in a market with hidden orders minimizes the expected losses of the uninformed.

When the book is displayed, liquidity demanders can infer some of the private information possessed by liquidity providers. This expropriates informational rents from the informed who do trade as liquidity providers, which in turn deters some of them from providing liquidity. Informed trader participation shifts away from the liquidity-provision side toward the liquidity-demand side of the market (i.e., $J^* < M$) as depicted in Figure 2 in the paper. This shift reduces competition overall resulting in greater expected losses to the uninformed and less informationally efficient midquotes than if all the informed were to trade as liquidity providers. Consequently, market quality is worse in the market with displayed orders than in the market with hidden orders.

Rents to providing liquidity draw the informed into trading aggressively. Because rents tend to disappear as markets grow, one might expect the difference in market quality between the hidden and displayed market structures to disappear as the market grows—i.e., as $M \to \infty$. However, the differences persist because, as the market grows, traders adjust their trading intensity to account for the size of the market. Display continues to provide a deterrent to liquidity provision, and the differences between hidden and displayed markets remain even when M is large.

The formulas that describe large markets turn out to be quite simple. When orders that provide liquidity are displayed, $J^* = M - M^{1/3}$; whereas $J^* = M$ when orders are hidden. This means display deters $M^{1/3}$ of the population of informed traders from providing liquidity even when the market is large.

Display deters liquidity even when the market is large

This shift away from liquidity provision affects the intensity of competition even though the market is large. Informed traders who provide liquidity in a displayed market submit orders that are only a fraction of the quantity that are submitted in a market with hidden orders. This fraction is approximately equal to $1/(M - M^{1/3})$ in a large market. For example, when $M = 100$, the informed submit orders to a displayed book that are only 1/95th as aggressive as the orders they would submit to a hidden book.

The impact on market quality is simple to characterize as well. As the market grows large, the magnitude of the expected loss of the uninformed traders grows in both markets, and remains greater by a factor of $2 \cdot M^{1/6}$ in the displayed market. The uninformed losses arise from transaction prices that deviate from fundamental value. Even in large markets, transaction prices depart more from fundamental value when orders are displayed than when orders are hidden.

Similarly, display discourages the incorporation of information into the midquote through the orders submitted by the informed into the book even when the market is large. We characterize this by the squared correlation (R^2) between the market-clearing price and order flow. In a large hidden market, half of the variability in prices is attributable to information impounded into the price schedule by informed orders that provide liquidity. The other half is attributable to the noise from uninformed trading that affects prices through order flow, and $R^2 = 0.5$. In a large displayed market, the aggressiveness of the orders submitted by informed liquidity providers decreases so much that the variability in prices attributable to these orders vanishes. In this case, all the variation in prices is attributable to the informed who trade as liquidity demanders and the noise generated by uninformed trading, and $R^2 = 1$.

The paper closes by demonstrating that the relevance of whether orders are hidden or displayed relies on traders having heterogeneous information. If all the informed observe the same signal, then the orders that provide liquidity are so aggressive that the common signal is fully priced into the order book. In this case, the informed derive no expected profit from their information even when orders are hidden, so there is no deterrent associated with display. The informed still earn expected profit from their role as liquidity providers. However, the equilibrium strategies of

the informed and measures of market quality are unaffected by whether information is hidden or displayed.

Bellini (1430–1516): Madonna of the Meadow. Italy, 1500. Can you imagine a highly reputed drawing and painting workshop in Venice in the 15th century, where the apprentices were busy churning out images of Madonna, under the direction of none other than Bellini? Bellini is considered to be the father of Renaissance painting, creating brilliant shades with the newly-perfected art of oil painting. The Madonna seems psychologically distant from the baby Jesus. This painting hangs in the National Art Gallery in London.

Lauren Cohen, Christopher Malloy, and Lukasz Pomorski

Uncovering the hidden information in insider trading

Corporate insiders have, by definition, considerably more information about their companies than what is publicly available. Their trades are closely followed by investors and the general public, who hope to glean from them new information about an insider's company and its future share price. But are all insider trades equal in their informational content? For example, suppose that you learned that Bill Gates—a savvy and undoubtedly well informed insider—sold 20 million shares of Microsoft in the third quarter of 2008. How would you interpret this bit of data? Did Gates anticipate the brewing crisis and sell his shares ahead of it? Or did he have some privileged information about Microsoft's future? Crucially, could one systematically make money by replicating his trades?

Whatever is your prior on Gates' motives, your evaluation of his actions would probably change when you found out that he sold another 20 million shares in the last quarter of 2008. And another 20 million in the first quarter of 2009. In fact, Bill Gates routinely sold 20 million shares in each subsequent quarter, as seen in Figure 1. It seems very unlikely that this pattern of trades could arise for information reasons. This means that if you wanted to use insider trades to learn about a company's future, you would probably want to ignore the trades of insiders who, like Bill Gates, are likely trading for reasons that have little to do with private information.

Figure 1: Gates' shares sold in MSFT each quarter (in million dollars)
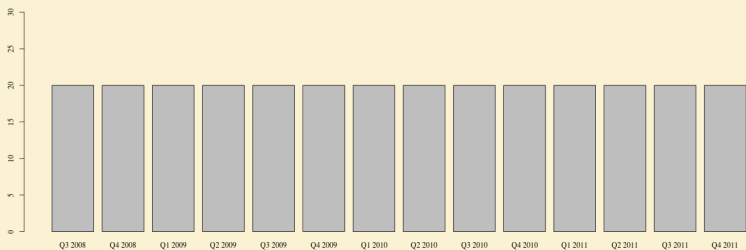
Are insider buys more informative than insider sells?

The Gates example serves well to illustrate the typical view of insider trading in the academic community. It has been well documented that while insider buys help predict future stock prices, there is little evidence of any price changes following insider sells. For example, Jeng-Metrick--Zeckhauser (REStat 2003) show that controlling for the risk exposures, stock prices go up by over 6% on average in the year following an insider purchase, but remain flat after an insider sale. The usual explanation of such evidence is that insider buys are indeed motivated by information, but that sells are made for other reasons. Insiders such as Bill Gates often have substantial stock holdings in their companies, and thus may want to sell some shares to diversify their overall portfolio. Moreover, insiders may sell their shares because of a specific liquidity need, e.g., they may be buying a house. Because such sells are not based on any firm-specific information, they should not be expected to predict future returns.

This binary view of insider trading (informative buys, uninformative sells) holds on average, but likely masks interesting variation. After all, some of the best-known examples of insider trading feature insiders selling on information: Enron executives liquidating their holdings ahead of their firm's bankruptcy, the selling of ImClone stock that ultimately led to the imprisonment of Samuel Waksal and Martha Stewart, etc. This means that at least some insider sells may be informative. Conversely, some insider purchases may not be based on new information. For example, some companies have stock purchase programs that allow insiders to purchase their company stock at a discount. Insiders who have money to invest (e.g., if they have just received their annual bonus) may want to participate in such programs even though they have no specific private information that would otherwise justify a trade.

The key question, of course, is how to distinguish trades that are likely to be based on information from trades motivated by other considerations. In Cohen-Malloy-Pomorski (JF 2012) we propose a simple way to divide trades into "routine," or less likely to be information-based, and "opportunistic," or more likely to carry new information. We show that our classification scheme works well in predicting company returns and news. Interestingly, we also find evidence that insiders limit their opportunistic trading following waves of SEC insider trading enforcement. In this article, we review our approach and our main findings.

We base our identification of opportunistic and routine insiders on the idea that trades based on private information are unlikely to follow predictable calendar patterns. This is the same argument we made earlier when discussing Bill Gates's trades: new information is unlikely to be generated in a regular calendar cycle. So, insiders who trade on information are likely to trade in a more irregular fashion. To test this idea we need to decide which calendar regularities to look for. In our paper we chose a very simple approach: we check whether a given insider trades in the same calendar month year after year. Insiders who trade in the same month in three consecutive years are classified as "routine," or unlikely to trade on information. Insiders who trade at least once per year for three years in a row but do not have a monthly routine are classified as "opportunistic," or more likely to act on information. Of course, this simple idea can be easily extended and perhaps refined.

What we show is that even this simple identification is enough to gainfully separate informative from relatively less informative trades.

Example from our sample

To better understand our approach, consider the following example from our sample. Electronics Corporation is a large, publicly traded firm, founded in the late 1960s. (The name of the firm and the dates involved have been disguised.) The firm had a number of insiders from 2003–2006. In particular, two of these insiders were actively trading, but in very different ways. The first insider (the routine trader) traded consistently over the time period, routinely trading in each and every March and solely in March. The second insider (the opportunistic trader), who also happened to be the CFO of Electronics Corporation, traded much differently. His trades came at very selective times over the same time period. As can be seen in Table 1, both employees traded 4 times over the 4 years. Further, their trades contained very different information for future prices. As shown in Table 1, the average returns in the month following the routine trader's sells were positive 33 basis points per month (so a –33 basis point return, as the insider was selling each March). In contrast, the average return following the opportunistic insider's sells was –5.69% per month (so a positive 5.69%).

What is important to note here is that both the opportunistic and routine traders were trading in their respective manners throughout their entire trading histories, so one could have predictably identified these traders as either opportunistic or routine traders before the period we have shown here. We exploit this ability to predictably classify insiders into these two classes of traders throughout the universe of traders.

Opportunistic trades move markets

Cohen-Malloy-Pomorski (JF 2012) show that across the entire universe of insiders from 1989-2007, a portfolio that mimics opportunistic insider trades—long a portfolio of opportunistic buys and short a portfolio of opportunistic sells—in the month following their trades (such that the portfolio is tradable), makes excess returns of 108 points per month (roughly 13% per year). The identical portfolio mimicking the trades of routine insiders earns only 27 basis points per month. The 81 basis point difference between the two (opportunistic–routine) is highly significant, and isolates the extra information in opportunistic trades for firm's future price movements over that of routine insider trades. This can be seen explicitly by rearranging the spread portfolio as (opportunistic buys–routine buys) = the extra information in buys; and equivalently (opportunistic sells–routine sells) = the extra information in opportunistic sells. The helpful aspect of this decomposition is that it's easily seen that if one did not distinguish between types of insiders, viewing all insiders as informed traders trading on that information, then one would be missing the rich information that emerges from stripping away routine trades.

Table 1: Returns in months following insider sales at Electronics Corporation

	Routine Trader	Opportunistic Trader
Avg Return in Month Following Sales	0.33%	–5.69%
# of trades	4	4
Dates of Trades	Mar-03, Mar-04	Jun-03, Apr-04
	Mar-05, Mar-06	Jul-05, Nov-06

In Table 2, we update this data through 2011. The same pattern emerges. While both returns increase, the difference between the two (opportunistic–routine) is 85 basis points over this most recent 4-year period. We include each leg of the trade to show the incremental information of opportunistic traders over routine traders, separately for insider buying and selling.

Table 2: Monthly value-weighted portfolio returns, averages in percent

		2008–2011	2009–2011
Buys	Opportunistic	1.40	2.91
	Routine	0.91	1.98
	Difference	0.49	0.93
Sells	Opportunistic	–0.35	0.70
	Routine	0.01	1.13
	Difference	–0.36	–0.43
Buy – Sells	Opportunistic	1.75	2.21
	Routine	0.90	0.85
	Difference	**0.85**	**1.37**

Routine trades showed much less of a stock-price response than opportunistic trades. The difference actually widened in the second half of the sample. 1.37% per month was both statistically and economically significant.

Summary

The routine-opportunistic classification we have proposed is simple and intuitive. It is also easily extended. One could, for example, use a more complicated pattern in trades to define "routines" or perhaps use additional data to improve the classification. The approach we propose also lends itself to applications in other context, such as routine rebalancing trades that might be done by institutional or pension managers each quarter.

Our work has attracted interest from policymakers, e.g., the SEC and the Ontario Securities Commission. It is also useful for market participants. For instance, Alliance Bernstein Research 2012 has produced a report that discusses, replicates, and confirms our results along with offering a number of extensions of ways that they may be useful in investment practice.
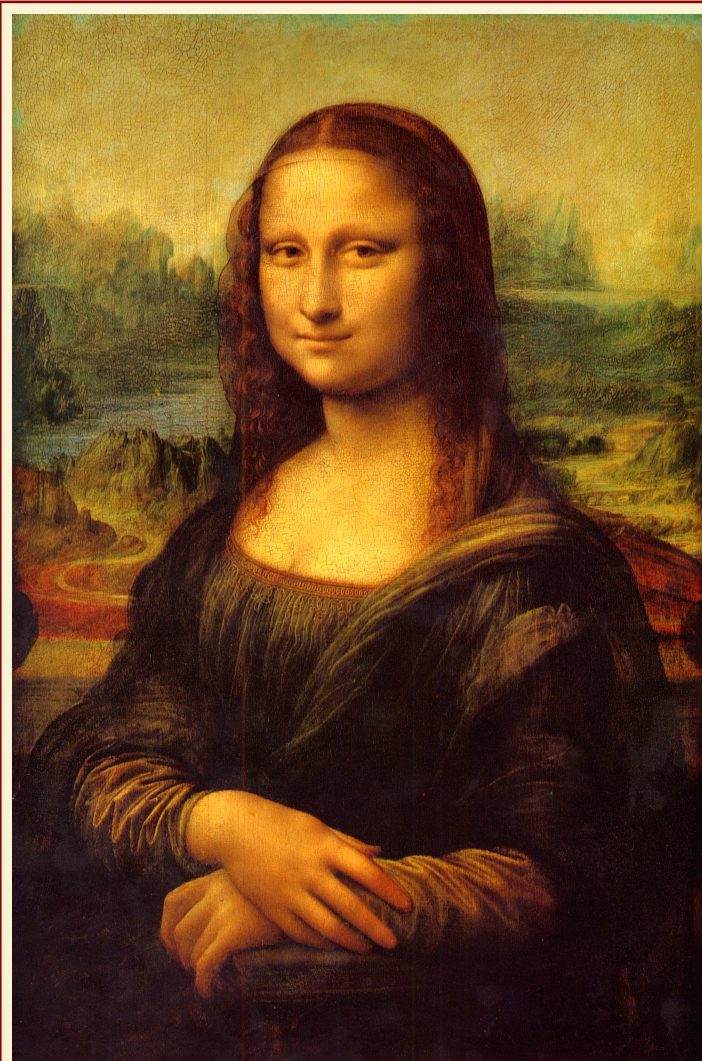
Leonardo da Vinci (1452–1519): Mona Lisa. Italy, 1505. Almost forgotten, this most-famous smile in the world was once stolen from the Louvre Museum in Paris. The patriotic Italian who effortlessly stole it in 1911 from a complacent Louvre just wanted it to be back in Italy—to restore a treasure stolen by Napoleon. Alas, the painting was in fact acquired by King Francis I of France and not by Napoleon. The director of Uffizi in Florence thus disagreed with this personal national restoration plan and returned the painting to the Louvre. The patriotic thief was stored, too.

The average investor buys stocks with the hope and anticipation of share price increases and thus profits. Yet, more sophisticated investors often engage in short selling which allows them to profit when stock prices fall. In short sales or short selling, investors borrow shares of a company from a stockbroker or other shareholder with the promise to return these shares at some point in the future. The borrowing investor immediately sells the shares in the open market and deposits the proceeds in their account. If the share price drops, the borrowing investor repurchases the shares at a lower price, returns the borrowed shares to their broker, and then realizes a profit.

Why would an investor engage in a short-sales transaction? Perhaps the investor has done his homework through a well-executed fundamental analysis and strongly believes that share price will decrease and thus makes a rational and reasonable "bet." For instance, an investor may have taken a short position in Blackberry and Nokia after Apple's introduction of the iPhone in 2007, recognizing that the new technology may substantially erode market share from these former industry leaders. Short sales however, also have a dark side. Investors holding negative, inside information about the firm's future prospects can use this knowledge to their private benefit and make almost assured profits at the expense of uninformed outside shareholders.

Insiders in family-owned firms may have more information

Our study focuses on the dark side of short selling and asks whether corporate ownership structure (who owns the firm's shares) leads to differences in informed short sales. Specifically, we examine short-sale activity in firms controlled by founding-family shareholders. Founding families hold large, undiversified ownership positions in about one-third of the S&P 500 firms with an average ownership stake of nearly 25% of the firm's shares and with family members holding the CEO post in about 50% of these firms. Family owners represent a select investor group with substantial control over firm activities and with privileged access to the firm's private information.

Family shareholders—founders and the founder's descendants or heirs—arguably possess strong incentives to engage in short-selling activity. These incentives may arise from their ability to profit from their access to private information. Alternatively, conflicts of interest among family members can lead those not employed by the firm to take destructive or harmful actions. Further, employees

who are not members of the family group may also be disgruntled with family interference or family domination of senior managerial posts, leading to a leakage of private information on firm activities. The family's close ties with the firm may generate a variety of linkages and incentives that could facilitate the use and dissemination of private, negative information on corporate activities that could lead to an increase in short selling.

Do family insiders seem to sell based on information?

Although several viable arguments exist to suggest that family firms experience more informed short selling than non-family firms, anecdotal accounts suggest that family shareholders have strong incentives and mechanisms to limit informed short sales in their firm's shares. Patrick Byrne, the founder of Overstock.com, for instance, attempted to use the courts to limit short sales in his firm's shares, arguing that such activities by hedge funds were increasing the firm's cost of capital. Family owners can also limit informed short sales by asking shareholders to move their shares from margin accounts into personal accounts or withholding the family's own substantial stakes from circulation, thereby increasing borrowing costs for short sellers. Overall, these powerful shareholders have strong incentives to protect the family's reputation, limit public visibility, and safeguard wealth, potentially deterring informed short sales in their firm's shares. We seek to answer the question of whether family presence facilitates or hinders informed short selling.

To answer this question, we examine daily short sales in publicly-traded U.S. firms between January 2005 and July 2007. Our testing procedures follow two distinct paths. First, we focus on negative, quarterly earnings surprises. That is, quarterly earnings where the firm failed to meet the pre-established benchmark set by the market and stock analysts. Do family firms (relative to non-family firms) systematically experience greater short selling in advance of these unexpected negative-earnings surprises? If so, this would be evidence suggesting that select investors receive and act upon negative information prior to its release to the general market. Second, we focus on whether the volume of short sales successfully allows investors to predict future stock returns in family firms versus non-family firms—irrespective of any information event. Using today's short-sales volume, can an investor in family firms systematically improve future returns more than an investor in non-family firms or more specifically, "beat the market"? If investors in family firms can routinely beat future market returns by using today's short-sale volume, then this would also be indicative of select investors receiving and acting upon negative, confidential information prior to its release to the general market.

Our sample consists of 1,571 largest U.S. industrial firms as of January 2004, with daily short sales data (from SEC REG SHO database) available during January 2005 to July 2007. In the quarterly tests, we examine abnormal short sales prior to negative quarterly earnings surprises and, for symmetry, prior to positive quarterly earnings surprises. The 1,571 firms in our sample provide 4,702 negative quarterly earnings surprises and 5,491 positive quarterly earnings surprises from January 2005 to July 2007. The daily tests examine whether short-sale interest (volume) on day t predicts future stock returns on day $t + 2$. The 1,571 firms in our sample provide 310,720 firm-day observations for family firms and 523,264 firm-day observations for non-family firms from January

2005 to July 2007. Based on a minimum 5% ownership threshold, family firms constitute 36% of the sample with an average ownership stake of 25.3%.

Yes, they do!

For the quarterly earnings announcement tests, we calculate abnormal short-sale volume as [(average daily short sales during the event window divided by average daily short sales for the year outside of event window) minus 1]. The event window for calculation is the 30-days prior to quarterly earnings announcements. Our analysis indicates that short sales in family firms are over six times more sensitive to the magnitude of future negative earnings shocks than short sales in non-family firms. Further, family firms experience almost seventeen times more short selling preceding negative earnings shocks than non-family firms, suggesting extensive informed trading. In real-world terms, if our sample firms announce earnings that are $0.10 lower than expected, our results indicate that short sales in family firms, on average, increase by about 11,433 shares for each of the 22 trading days before the announcement date. In contrast, non-family firms experience an increase in short-sales volume of about 683 shares each day. Our analysis provides strong support to the notion that select investors in family firms systematically receive and act upon negative information prior to its release to the general market; consistent with an informed trading explanation.

In our second set of tests, we relax the assumption of a negative information event (i.e., quarterly earnings announcements) and examine whether an investor can systematically beat the market by using today's short-sales volume to predict future stock returns. We use the Fama-French three- and four- factor models as our benchmark measure of future stock returns. Our analysis indicates that an investor who routinely shorts family firms with the greatest volume of short sales and buys family firms with the lowest volume of shorts sales will outperform the market by 60.9 basis points per month or 7.31% per year. A similar strategy for non-family firms yields no profit. Our results from the daily stock return tests again provide compelling evidence that select investors in family firms systematically receive and act upon negative information prior to its release to the general market.

In the study, we experience a data limitation in that we only know the volume of short selling for each firm but do not know the identity of individual short sellers. These sellers could be family members, managers, distant relatives, or well-informed outside investors. To provide further insights into the identity of the short seller, we examine whether short sales volume is related to the ownership of hedge funds, private equity funds, mutual funds, pension funds and insurance companies. The analysis suggests that ownership by these large influential shareholders does not generally affect the volume of short sales prior to negative, quarterly earnings surprises. In additional testing however, we find a substantial increase in short-sales volume when a family member (founder or descendant) serves as CEO, when family members control two or more board seats, and when families use dual-class share structures. The evidence indicates that as family control increases, short-sales volume increases prior to negative corporate events.

Regulatory scrutiny—or lack thereof

Even though informed trading can facilitate corporate transparency and help improve the efficiency of the stock market, U.S. regulators seek to restrict individuals from trading on material, non-public information. In 1984, the U.S. Supreme Court ruled that trading based on information received through breaches of fiduciary responsibilities, such as managers providing updates to family shareholders so that they can trade on the information, represents illegal insider activity. Given that our evidence indicates significant informed trading among family firms, a natural question is how much of this activity is pursued or detected by regulators. Interestingly, our search of Security and Exchange Commission (SEC) enforcement reports shows that the Commission extensively targets hedge funds for insider trading but appears to have little focus on family firms. For instance, from 2006 to 2008, twenty-two enforcement actions were brought against hedge funds and none against family owners. As such, our analysis adds to the question of the general efficacy of insider trading regulations in limiting such activity by informed traders. Our results suggest that regulations may be effective in reducing informed trading in non-family firms but appear substantially less effective in family firms. There are two possible explanations.

First, the scope or definition of insiders and insider trading in the U.S. may not be appropriate. The U.S. Congress and the SEC can impose severe monetary penalties for insider trading but do not explicitly define this activity in legislation. Rather, Congress and the SEC typically allow courts to define the scope of insider trading, which requires a breach of fiduciary responsibility or duty for trades to be considered illegal. In other words, U.S. courts have held that trading based upon non-public information obtained without a breach of fiduciary responsibility is legal. The murkiness of the insider trading definition implies that individual shareholders with less than 10% ownership stakes can trade on material non-public information as long as the information is obtained without a breach of trust. Family owners often have their ownership dispersed amongst several different family members and/or trusts, such that no single individual or entity holds more than 10% of the firm's shares.

Second, U.S. regulations provide more severe penalties for trading based on private information regarding mergers and acquisitions relative to routine earnings announcements (e.g., our measure of negative corporate events). Regulators explicitly justify the differential treatment by arguing that long-term shareholders bear relatively little harm (or benefit) from short-term stock price fluctuations around earnings announcements. Our study clearly indicates profitable short selling based on routine earnings announcements, consistent with the notion that short sellers can use quarterly earnings announcements as a relatively obscure and arguably safe route to engage in informed trading. Overall, our study has strong policy implications and suggests that (i) regulators and government agencies may want to expend greater resources on detecting insider trading in family firms; (ii) lawmakers may want to develop a more inclusive definition of corporate insiders who can be held liable for misappropriation of confidential information with explicit consideration of family members; and (iii) law enforcement agencies may need to focus on less significant corporate events in detecting insider trading versus simply concentrating on large corporate transactions.

Michelangelo (1475–1564): David Statue. Italy, 1504. This 17-foot statue of shepherd boy David stands tall in the Gallery Academia in Florence. To many Italians, it symbolizes liberty and freedom. The self-assured renaissance man David represented the real way of worshipping God—not spending hours in prayers, but recognizing and using human talents. By the way, Michelangelo—often considered to be the greatest visual artist ever—created the David after studying the inside of human corpses. The quality of the marble is inferior, though, meaning that there is a good chance that it will not last forever.

Time series and cross-sectional average returns are computed for a wide variety of purposes in investment and financial analysis. Further, it is common to estimate conditional (on the outcomes of various explanatory variables) mean returns by implementing regression analysis. The computation and comparison of mean rates of return for securities and portfolios may well be the single most common empirical method used in the field of Finance.

In our paper "Noisy Prices and Inference Regarding Returns" we show that these simple calculations can give misleading results if prices contain "noise," by which we mean temporary deviations of observed prices from underlying security values. Noise can arise from microstructure frictions, such as the temporary price impact of large orders. Noise can also enter prices due to traders' behavioral biases, if arbitrage is imperfect.

Noisy prices bias average returns upwards

In particular, the simple mean return based on noisy prices provides an upward biased estimate of the rate at which the true values and observed prices trend upward (e.g., Blume-Stambaugh (JFE 1983)). A simple example illustrates the issue. Suppose that fundamental value is constant at \$10, implying that the mean true return (i.e., that computed from fundamental values) is zero. Due to market imperfections, trades occur at noisy prices of \$9 or \$11, equally likely. The possible returns computed from trade prices are 22.22% (if the price moves from \$9 to \$11), –18.18% (if the price moves from \$11 to \$9), or zero (if the price moves remains at \$9 or \$11). In a large sample, the mean observed return will be 1.35%, even though value and prices are not trending upward at all! If this example were repeated with more dispersion attributable to noise in prices, the mean observed return would be greater, and vice versa. (Our observations regarding the effect of noisy prices are related to, but distinct from, the well-known relation between arithmetic and geometric mean returns. The arithmetic mean return exceeds the geometric mean when there is any variability in returns. The preceding statement holds for both true and observed returns. In contrast, the bias between the (arithmetic) mean observed and true returns arises only if prices contain temporary deviations from value.)

Importantly, the bias in returns is always positive when prices contain noise (even though the noise itself is zero mean), implying that the bias is not diversified in portfolios. (Simple means of observed returns do have an economic interpretation. In particular, they give the outcome to a hypothetical active trading strategy that involves selling securities that have appreciated and purchasing securities that have depreciated, so as to reestablish equal weights, while assuming that the transactions can be completed at observed prices. However, because only a subset of traders can engage in any active trading strategy, the returns to this hypothetical rebalancing strategy are generally not the same as (and if prices contain noise on average exceed) the returns to investors in aggregate.) If securities are sorted into portfolios based on an attribute correlated with the amount of noise (e.g. firm size, illiquidity, or volatility) then the mean return to the portfolio containing noisier securities is upward biased by a greater amount. In an earlier paper (Asparouhova-Bessembinder-Kalcheva (JFE 2010)), we showed that the problem of noisy prices pertains not just to average return computations-it also leads to biases in coefficients estimated in any ordinary least squares (OLS) regression with returns as the dependent variable.

What can/should we do to adjust?

We consider two key issues. First, we assess three methods to correct for the bias, each of which amounts to computing weighted average returns. The methods considered are (i) weight each return by prior-period market capitalization (value weight, or VW), (ii) weight each return by the accumulated gross return from a specified formation period until the end of the prior period (initial-equal-weight, or IEW), or (iii) weight each return by the gross return in the prior period (return weight, RW). We show by theory and simulation that each is effective in removing the biases attributable to noisy prices, under reasonable assumptions. (While all three methods are effective in removing bias, their interpretation differs, because value-weighted returns place more emphasis on large firms, which can dominate a sample. A researchers preference for VW vs. RW or IEW corrections may depend on the relative importance of the information contained in large vs. small sample securities. Also, because it is common to form portfolios on an annual basis, we consider the effect of weighting by firm value measured as of the prior December. We show that the resulting mean returns remain upward biased.) Each is effective because it weights observed returns by a variable proportional to the prior-period observed price. If this price contains positive noise then the weight increases exactly as the observed return for the current period decreases (and vice versa when the noise is negative), resulting in weights and observed returns counteracting each other to offset the original upward bias in returns.

Table 1: Mean returns to attribute-sorted portfolios, Jan 1966 to Dec 2009

Size	Dec 10	Dec 1	Dec 10–1	(Std.Err.)
EW	0.462	1.888	−1.425***	(0.32)
RW	0.448	1.410	−0.961***	(0.31)
IEW	0.450	1.194	−0.743***	(0.31)
VW	0.371	0.888	−0.517**	(0.30)

Book-To-Market	Dec 10	Dec 1	Dec 10–1	(Std.Err.)
EW	1.517	0.148	1.369***	(0.23)
RW	1.301	0.024	1.277***	(0.22)
IEW	1.331	0.033	1.298***	(0.22)
VW	1.031	0.214	0.816***	(0.26)

Inverse Price	Dec 10	Dec 1	Dec 10–1	(Std.Err.)
EW	1.832	0.579	1.252***	(0.38)
RW	1.214	0.569	0.645	(0.37)
IEW	1.035	0.583	0.452	(0.36)
VW	0.570	0.405	0.164	(0.41)

Volume	Dec 10	Dec 1	Dec 10–1	(Std.Err.)
EW	0.407	1.605	−1.198***	(0.27)
RW	0.396	1.243	−0.846***	(0.26)
IEW	0.414	1.098	−0.684***	(0.25)
VW	0.354	0.713	−0.359**	(0.21)

Illiquidity	Dec 10	Dec 1	Dec 10–1	(Std.Err.)
EW	1.580	0.441	1.139***	(0.29)
RW	1.211	0.430	0.780***	(0.28)
IEW	1.108	0.433	0.675**	(0.28)
VW	0.769	0.356	0.413	(0.29)

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

The table reports time-series means of monthly returns to the extreme of the 10 attribute-sorted portfolios and to the corresponding hedge portfolio. Portfolio returns for month t are measured on an equal-weighted (EW), return-weighted (RW, weight is period t – 1 gross return), equal-initial-weighted (IEW, weight is cumulative gross return from portfolio formation through month t – 1), and prior-month-value-weighted (VW, weight is month t – 1 market capitalization). Firms are assigned to portfolios based on attributes measured in July.

Second, we assess how important the biases attributable to noisy prices are in real world data. Table 1 shows mean monthly returns to portfolios of U.S. equities from 1966 to 2009. We report the upward biased equally-weighted (EW) returns, as well as returns corrected for bias by use of the RW, IEW, and VW methods. The portfolios are created by sorting on firm characteristics, including firm size, book-to-market ratio, (inverse) share price, trading volume, and stock illiquidity. For each attribute, we sort stocks into ten portfolios, and report returns to the extreme (first and tenth) portfolios, as well as to the "hedge portfolio" that is long the tenth and short the first decile portfolios. Hedge portfolio returns are often interpreted as the mean return premium associated with investing in stocks with greater levels of the sorting characteristic.

The key observation to be gleaned from Table 1 is that the absolute value of the estimated return premium associated with each of these characteristics is larger when focusing on the (biased) equal-weighted (EW) returns than when considering the corrected estimates obtained by any of the IEW, RW, or VW methods. The bias is quite relevant for firm size, share price, trading volume, and illiquidity, but less so for the book-to-market ratio. Focusing in particular on the differential between EW and RW mean returns, noise in prices explains about one third of the apparent size effect in monthly returns, as the corrected size premium is –0.96% per month, as compared to the uncorrected estimate of –1.43% per month.

Results regarding the relation between mean returns and (inverse) share price are particularly striking, as they illustrate that inference regarding the existence, and not just the magnitude, of a return premium can be altered by bias attributable to noisy prices. Here, the estimated return premium is 1.25% per month based on EW returns, versus 0.65% per month (RW), 0.45% (IEW), or 0.16% (VW). None of the corrected estimates of the return premium associated with share price is statistically significant. That is, the biased (EW) estimate implies the existence of a return premium associated with share price, but none of the bias-corrected estimates do.

The importance of the bias in mean returns depends on how noisy the prices are. Some researchers exclude low-priced securities (e.g. those trading below $5) from their studies, reasoning that these will be most affected by noise. We show that excluding low priced securities is indeed quite effective in reducing the bias. However, statistically significant bias persists. More importantly, excluding noisy securities from a study reduces statistical power and discards important information regarding the magnitude and form (e.g. linear in firm characteristics) of return premia.

Conclusion

In summary, our study considers five specific firm characteristics, using monthly return data for U.S. equities since 1966, and we document significant biases in return premia estimates obtained by comparing EW returns across attribute-sorted portfolios, as well as estimates obtained by cross-sectional OLS "Fama-MacBeth" regressions. We anticipate that many of the return premium estimates that have been reported in the literature as being associated with additional firm characteristics and/or with factor exposures are also biased. We caution that biases are likely to be greater in some other applications, e.g. in studies of mean returns to corporate bonds or international equities, if prices in those markets contain more noise. Further, biases of the type we study are

likely to be relatively more important in daily or higher-frequency returns, because true returns are smaller at shorter horizons, while the biases attributable to noise are not.
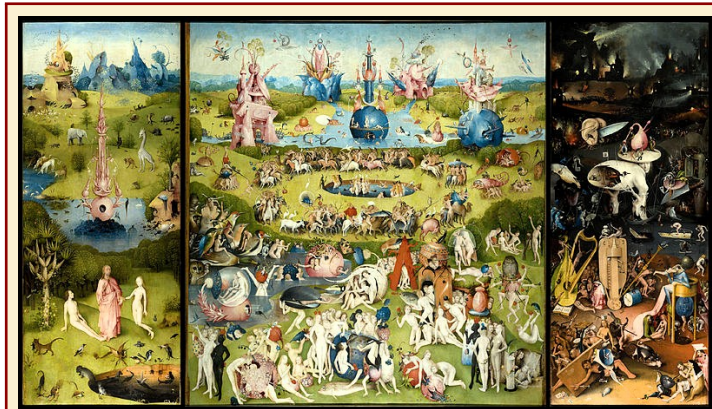
Hieronymus Bosch (1450–1516): The Garden of Earthly Delights. Flemish Renaissance, 1505. Bosch probably meant for this complex painting as a caution against hazards of sinful delights. What Bosch could not imagine was that a music student from Oklahoma would transcribe music she found in this work, "written upon the posterior [butt] of one of the many tortured denizens of the rightmost panel of the painting" into modern notation. She calls it the "600 year old butt song from hell."

In this paper, we study the debt maturity in US industrial firms from 1976 to 2008. We find a sharp decrease in debt maturity, with the median percentage of debt maturing in more than 3 years decreasing from 64% in 1976 to 49% in 2008. We take different approaches to investigate this trend: first we look at firm characteristics to check whether the main debt maturity theories can explain it; second we investigate whether the decrease in debt maturity is the result of demand-side factors or the result of changes that are not related to firm characteristics (supply-side effects); and finally we show the importance of new listings in explaining the trend.

Our data are taken from the Compustat Industrial Annual database. We exclude financial firms and utilities. The final sample has 12,938 unique firms with a total number of 97,215 observations. Our proxy for debt maturity is the percentage of debt maturing in more than 3 years (Barclay-Smith, JF 1995).

Debt maturities declined for decades

We identify a strong decrease in debt maturity over the sample period, in particular until the year 2000 when a reversal takes place. The decline is stronger for longer maturities. Leverage ratios remain stable over the same period, which suggests that the shift towards shorter maturities is not related to structural changes in the usage of leverage. The downward trend in debt maturity is statistically and economically significant: the median of the proportion of debt maturing in more than 3 years decreases 0.61% per year. We examine the time trend in debt maturity across firms of different sizes: small, medium-size and large firms. We find that debt maturity is significantly shorter for small firms, but also, that the decrease in debt maturity is much stronger in this group of firms. The number of firms and its evolution in the sample period is also much different between the small firms group and the other two (see Figures 1 and 2).

Figure 1: Number of firms
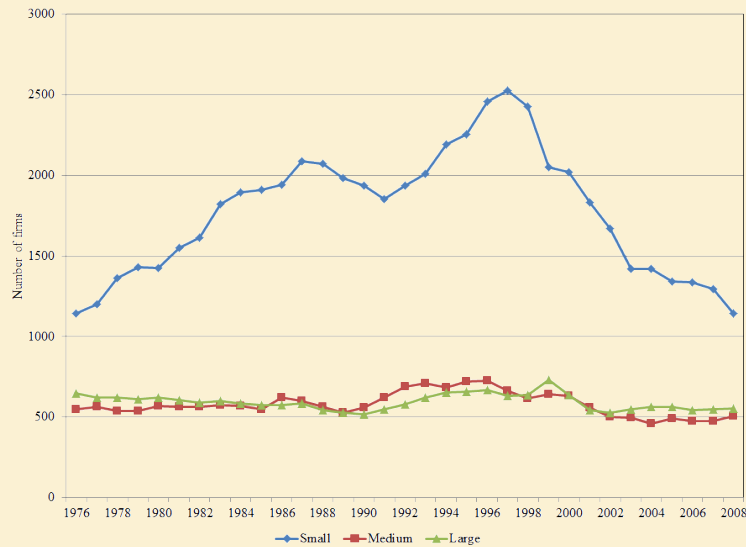
Figure 2: Percent of long-term debt maturing in 3 years or more
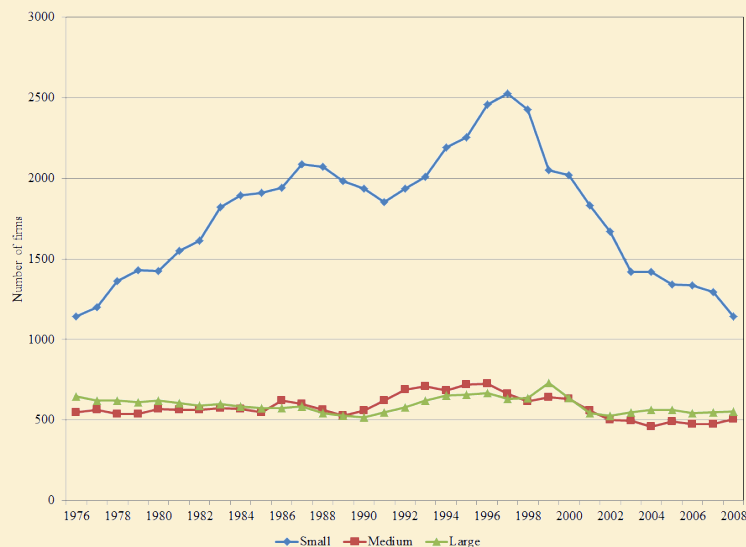
It's not agency, information, signaling, or liquidity

We then test whether debt maturity theories can explain the time trend. We find no evidence that agency costs of debt (resulting from conflict of interest between shareholders and debt holders), nor managerial agency costs (resulting from conflict of interest between shareholders and managers), explain the decline in debt maturity over time. Firms with lower agency costs of debt experience

significant decreases in debt maturity. When we categorize firms by proxies of managerial agency costs—governance index (Gompers-Ishii-Metrick (QJE 2003)) and managerial ownership—we do not see different patterns across groups of firms.

We then investigate the role of information asymmetry between the firm and shareholders (or different access to information). We use balance sheet proxies of information asymmetry (level of research and development and tangibility of assets), dynamic proxies (bond rating, institutional ownership, analyst coverage, dispersion of analyst forecasts, asset volatility) and a market microstructure measure of adverse selection (illiquidity measure of Amihud (JFM 2002)). We find evidence that firms with more information asymmetry use more short-term debt, which is consistent with the information asymmetry theory. More important to our analysis, we observe that firms with higher information asymmetry exhibit a decline in debt maturity.

The signaling (when firms try to signal the market their private information with the type of actions taken) and liquidity risk (the risk a firm with debt faces of not being able to refinance debt) theories do not offer any support in explaining the decrease in debt maturity. Contrary to the signaling theory, we find that firms with better projects have higher debt maturity than firms with worse quality projects (proxied by abnormal earnings or credit quality). We also find no distinct pattern in the evolution of the debt maturity between the two groups of firms.

It's younger firms

The decrease in debt maturity seems to be related to the disappearing dividends phenomena (Fama-French (JFE 2001)). Non-dividend payers use more short-term debt and decrease their debt maturity over time, as opposed to the dividend payers group of firms. Less profitable firms and firms holding more cash (Bates-Kahle-Stulz (JF 2009)) are also related to the negative trend in debt maturity.

We investigate whether the decrease in debt maturity is a result of demand-side factors, or a result of changes that are not related to firm characteristics using multivariate regression tests. We find that changes in firm characteristics explain part of the trend in debt maturity but they cannot fully explain it. Unobservable differences between firms and changes in the impact of firm characteristics on debt maturity also have limited power in explaining the evolution of debt maturity. Thus, firms are using more short-term debt, irrespective of their characteristics. Indeed, the expected debt maturity, generated by a regression model estimated using the earlier part of the sample period, systematically overestimates the actual maturity and consequently fails to fully capture the decrease in maturity.

We next show the importance of the firm listing cohort to explain the debt maturity trend. We categorize firms into four cohorts using the firm's listing year. Figure 3 shows the yearly evolution of the median debt maturity by listing groups.

Figure 3: Median debt maturity by listing group
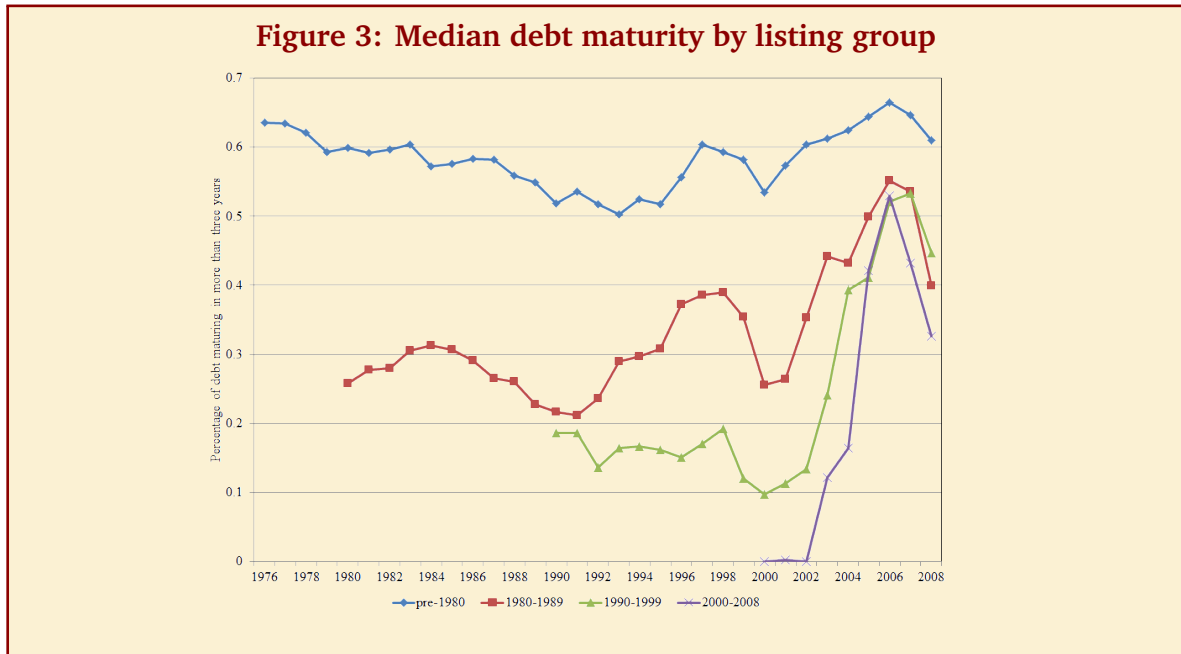
From Figure 3 it is clear that firms in the most recent listing groups use more short-term debt. We can also observe no negative time trend within each group. We conclude that the change in the sample composition of firms is a key factor to explain the decline in debt maturity. We perform additional tests to support this major finding: first we investigate if our findings are directly related to firm age, and conclude that the decline in debt maturity is not fully explained by firm age; next we test whether there is individual time trend for each firm with at least 5 and 10 yearly observations. The results show that the large majority of firms have an insignificant time trend, and that less than 20% have a negative and significant trend. Finally, we use a balanced panel (a panel where in all the years of the sample the firms are the same, thus excludes new listings) and we do not find a trend in the median debt maturity.

To further investigate the importance of the new listing in explaining the decrease in debt maturity, we run several panel regressions. When modeling debt maturity against a time trend and controlling for the listing decade (we use 4 dummy variables regarding firms listed in each of the 4 decades covered in our sample) we find that the trend coefficient (the average change on the debt maturity level from moving one year forward) is positive and statistically significant (from zero), which suggests that the average debt maturity increases through time. We also observe that, on average, firms listed on more recent decades have shorter debt maturity levels, with the exception of the last decade. When we add to our model firm characteristics, the time trend coefficient loses its explaining power. Controlling for the firm age and the founding age (following Jovanovic-Rousseau (AER 2001) and Loughran-Ritter (FM 2004)) keeps our main conclusions intact: there is no significant trend in debt maturity when accounting for the firm's listing year; young firms listed in the 1980s and 1990s use on average more short-term debt than a comparable firm listed in the 1970s. In summary, the new listing effect in necessary and sufficient to explain the negative trend in debt maturity.

We also look into industry composition and the evolution of debt maturity by industry. From a total of 49 industries (following Fama-French (JFE 1997) classification) we find 31 with a negative time trend, of which 23 are statistically significant (e.g., medical equipment and computer software). On the other hand, we find only one industry (petroleum and natural gas) with a positive and significant time trend. Industry composition also seems to play a role in explaining the decrease in debt maturity, because industries with stronger decreases in debt maturity also have stronger increases in market capitalization.

Using data from Worldscope for 23 developed countries for the 1990–2008 period, we check whether the decrease in debt maturity is exclusive to the US. Overall we find no evidence of a decrease in debt maturity outside of the US. When analyzing the data individually for four other major countries (UK, Germany, France and Japan) we only find a decrease in debt maturity in Japan. The Japanese trend can mostly be related to the existence of low levels of short-term interest rates during the sample period.

The supply of credit also mattered

Finally, we look at supply-side effects that might explain the trend in debt maturity. We begin by studying the maturity of new bond issues and syndicated loans. The sample of bond issues contains 12,821 issues for 1,986 industrial firms over the 1976–2008 period. We find a negative and strongly significant trend in the average and median maturity of new bond issues. The median maturity falls from 25 years in 1976 to 7.5 years in 2008. This trend is common to all size groups (including large firms), which differs from the results using balance sheet data. Using regression analysis, we find a negative and significant trend in the maturity of bond issues of all size groups, and also that this trend is not fully explained by the listing year. When examining the sample composition of bond issuers by size, we find that small firms have increased their importance in the public debt markets, which helps to explain the decrease in debt maturity.

To study the private debt market, we use a sample of 113,094 syndicated loans from 5,114 firms for the period 1987–2008. We find no evidence of a time trend. We also do not find a significant increase in the relative importance of small firms during the 1990s, as we do for bond markets. When we study the volume of private debt compared to public debt, we find an increase in the share of public debt. Together with the decrease in public debt maturity, this supports the idea that the decrease in debt maturity has taken place mainly in public debt markets.

To further investigate how the supply of credit affects debt maturity, we use an exogenous contraction in the supply of speculative-grade credit after 1989 resulting from regulatory changes derived from the collapse of Drexel Burnham Lambert. Speculative-grade firms significantly reduced their use of long-term debt relative to investment-grade firms. We also find that unrated firms decreased debt maturity after 2007 (i.e., at the time of a large contraction in the supply of bank loans) significantly more than rated firms. Thus, we conclude that debt maturity is affected by supply-side factors.

Conclusion

Overall, we find a secular decrease in the debt maturity, concentrated among small firms and firms with high information asymmetries. New firms listed in the 1980s and 1990s are responsible for the negative trend, because they use much more short-term debt than older firms. This trend is inexistent when we consider the firms listing vintage. However, we find that demand-side factors cannot fully explain the decrease in debt maturity. Supply-side factors are also relevant. The decrease in debt maturity occurred mainly in the public debt markets and has increased the exposure of firms to credit and liquidity shocks. This may have as well exacerbated the effects of the 2007–2008 financial crisis on the real economy.

Botticelli (1445–1510): La Primavera. Italy, early Renaissance, 1484. Unless you are an avid scholar of Greco-Roman mythology and the political interpretations of figures in this painting, you will overlook the point of this painting. It is supposed to depict a message from the Humanistic philosophy thriving in the 15th century Florence, which professed the human triumph of freedom over tyranny. Fortunately, it is nevertheless easy to appreciate the famous master-painter Botticelli's work in creating this beautiful garden of Venus. Alas, poor Botticelli died in obscurity, having been overshadowed by younger masters, such as Leonardo de Vinci, Raphael, and Michelangelo. It was only in the late 19th century that Botticelli was reinterpreted to have been among the earliest and finest old masters of his time.

Investing in private equity is among the preferred choices for long-term investors, such as endowments and pension funds, who seek to diversify their portfolios. These long term investors are clearly the best suited for holding an illiquid asset like private equity. The diversification benefits of private equity, however, have not been widely documented. In particular, an issue which has not been addressed so far, is whether private equity performance, like that of other asset classes, is affected by liquidity risk.

Liquidity risk has been studied by, among others, Pastor-Stambaugh (JPE 2003), Acharya-Pedersen (JFE 2005), and Sadka (JFE 2006), as an additional source of systematic risk for public equity. In general, liquidity risk refers to the co-movement between unexpected changes in overall market liquidity and asset returns. If an asset tends to have low returns when the overall market exhibits low liquidity, then this asset bears more systematic risk than an asset whose returns increase in low aggregate liquidity states. As a result, this asset's cost of capital should be higher.

A four-factor model for private equity returns

Prior studies have shown that liquidity risk is a relevant component of the cost of capital in different asset classes. Beyond public equity returns (e.g., Pastor-Stambaugh (JPE 2003), Acharya-Pedersen (JFE 2005), Sadka (JFE 2006), evidence of liquidity risk has been found for emerging markets, bond markets, credit derivative markets, and hedge funds. The primary goal of this paper is to quantify liquidity risk in private equity. To carry out this task, we use the four-factor model of Pastor-Stambaugh (JPE 2003). This asset pricing model contains three factors in addition to liquidity risk. The first factor captures market risk, as in the standard CAPM. The other two factors, introduced by Fama-French (JFE 1993), capture two important sources of return variation. The SMB factor is long in small stocks and short in big stocks, while the HML factor is long in high book-to-market (value) stocks and short in low book-to-market (growth) stocks. Empirical evidence shows that these two factors have earned significant returns and so do the securities that load positively on them. The literature is still debating whether these high returns are due to a risk premium or to market anomalies. We do not need to take a stand on the source of these premia as, for our purposes, what matters is to control for known sources of variation in asset returns. The goal is to filter out from a fund's performance the component of returns that originates from these

four replicable factors. Using this four-factor model, we can offer one of the first estimates of the cost of capital for private equity, an asset class which, according to some estimates, has reached $3 trillion of assets under management in 2012. Once a cost of capital is estimated, we can assess whether this asset class delivers outperformance (alpha) or not.

Table 1: Cash flows of a typical investment

Date (in years)	Cash Flows	Reinvested Div (at 5% / Sem)
0.0	−100	0
0.5	0	0
1.0	0	0
1.5	0	0
2.0	0	0
2.5	50	50
3.0	0	53
3.5	0	55
4.0	150	208

$$\text{MIRR} = (208/100)^{1/4} - 1 = 20\%$$

IRR = 21%

PV Div at 15% Disc	119
PV Div at 17% Disc	108

The table shows the cash flows of a representative investment. It lasts for four years, pays a final dividend equal to 1.5 times the original investment, and pays an intermediate dividend in year 2.5 which equals half of the initial investment. We show the computation of the modified IRR with a re-investment rate of 5% per semester. At the bottom of the table we report the present value of the dividends using two different discount rates.
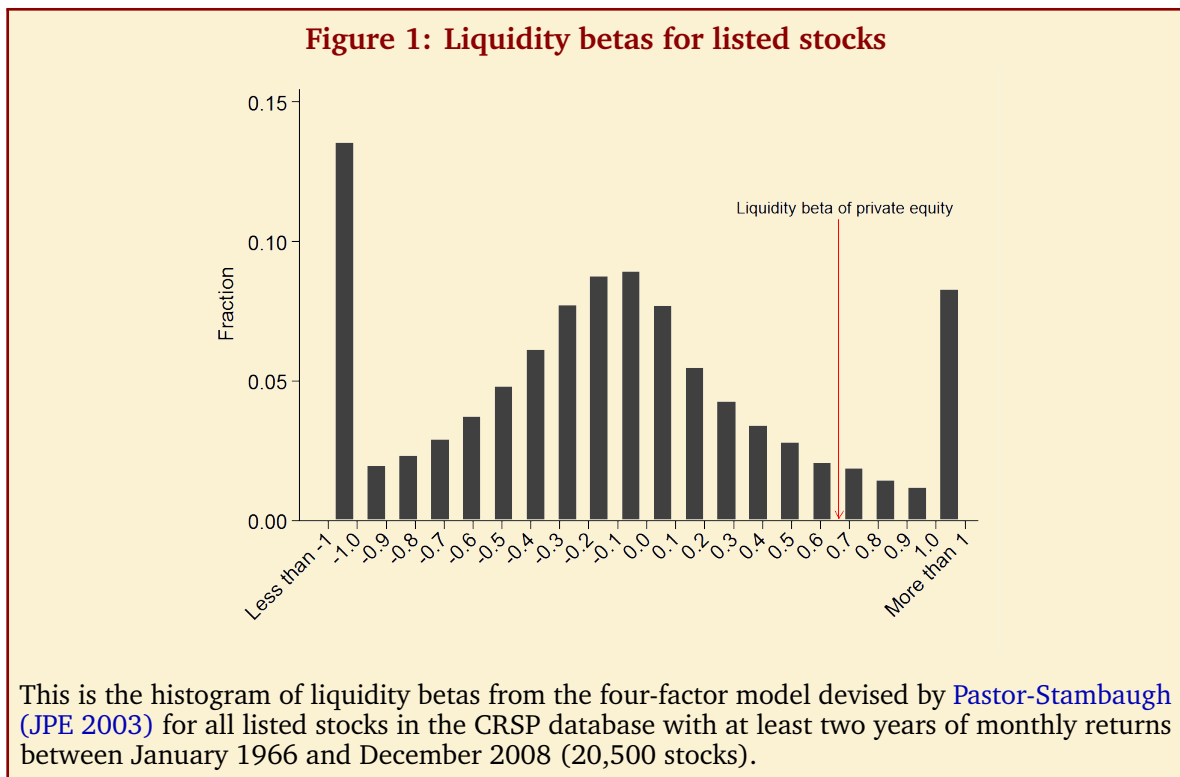

We use a unique and comprehensive dataset containing the precise cash flows generated by a large number of liquidated private equity investments. In order to clarify from the start the peculiar structure of our data, Table 1 shows a typical cash flow stream. (In our analysis we use a modified IRR (MIRR). It is like an IRR but instead of making the assumption that intermediate distributions are re-invested into the project at the IRR rate, it makes the assumption that the intermediate distributions are re-invested into the S&P 500. The reader may refer to Phalippou (JEP 2008) for a discussion of the use of IRR versus MIRR in private equity.) There is an initial negative cash flow (the investment) followed by two positive cash flows (an intermediate distribution and the final dividend corresponding to the divestment). Note that we do not have intermediate valuations

for the investment. So, there is no time series of returns, which precludes the use of the usual time-series regressions to estimate risk exposures. In such a context, as in Cochrane (JFE 2005), or Korteweg-Sorensen (RFS 2010), we exploit variation in returns across investments to estimate the risk loadings and abnormal performance of the asset class.

Accounting for liquidity risk, private equity has not outperformed

We fit the four-factor model of Pastor-Stambaugh (JPE 2003) to the data and find a significant beta on the liquidity risk factor (0.64), on the market factor (1.3), and on the value premium factor (1.0), but not on the size factor. These four factors together reduce the alpha of this asset class to zero.

Figure 1: Liquidity betas for listed stocks

This is the histogram of liquidity betas from the four-factor model devised by Pastor-Stambaugh (JPE 2003) for all listed stocks in the CRSP database with at least two years of monthly returns between January 1966 and December 2008 (20,500 stocks).

This historical 18% risk premium is significantly larger than the 8% hurdle rate commonly set in compensation contracts, suggesting that private equity investors (limited partners) were setting the bar too low for their fund managers (general partners). Expected risk premia may be lower as the equity premium seems to be declining (see, e.g. Graham-Harvey-Kolb (Book 2010)). Imagine a scenario in which the risk-free rate is zero, and each of the four risk premia is 3% p.a. then the benchmark will be: $1.3 \cdot 3\% + 1 \cdot 3\% + 0.64 \cdot 3\% \approx 9\%$. Under this scenario the 8% hurdle rate would make more sense. We feel that it is sensible to have time-varying hurdle rates as the risk premia are time-varying as well.

Importantly, the liquidity risk premium is about 3% annually, which implies a roughly 10% discount in the valuation of the typical investment (see Table 1). We also note that a liquidity risk

beta of 0.64 exceeds the corresponding estimate for the large majority (86%) of traded stocks. These results thus suggest that private equity is significantly exposed to the same liquidity risk factor as public equity and other asset classes. The diversification gains that can originate from private equity may thus be lower than previously thought given this additional risk exposure

Table 2: Explaining private equity returns (Log(1+MIRR))

Model	Market	FF	PS
IML			0.638***
			(0.180)
Rm – Rf	0.948***	1.395***	1.294***
	(0.14)	(0.26)	(0.25)
HML		0.719***	1.020***
		(0.39)	(0.29)
SMB		–0.124	–0.040
		(0.25)	(0.24)
Constant	0.006***	0.000	–0.002
	(0.001)	(0.000)	(–0.003)
Sigma	0.049	0.048	0.046
Adj. R^2	0.849	0.853	0.865
N, 1975–2007	139	139	139

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

Portfolios are formed by the starting date of the investment and must contain at least twenty investments. Each explanatory variable is computed by taking its average value during the portfolio life.Each observation is weighted by the square root of the investment duration to correct for unequal variance. The factor models are standard. IML is the Pastor-Stambaugh (JPE 2003) illiquid minus liquid portfolio.

Capital constraints and funding liquidity

Prompted by the finding of a significant loading on liquidity risk, we study the economic channel that relates private equity returns to market liquidity. We conjecture that due to their high leverage, private equity investments are sensitive to the capital constraints faced by the providers of debt to private equity, who are primarily banks and hedge funds. Adapting the Brunnermeier-Pedersen (RFS 2009) "funding liquidity" theory to private equity, the story is that times of low market liquidity

are likely to coincide with times when private equity managers may find it difficult to refinance their investments. In these periods, they may be forced to liquidate the investments or to accept higher borrowing costs, which in turn translate into lower returns for this asset class. Then, we conjecture that the link between private equity returns and market liquidity occurs via a so-called funding liquidity channel.

Empirically, we proxy for the evolution in funding liquidity with changes in the credit standards as reported in the Federal Reserve's Senior Loan Officer Survey. This survey asks loan officers at main banks whether they tightened or loosened their lending standards relative to the previous quarter. Axelson-Jenkinson-Stromberg-Weisbach (JF 2013) argue that, in the private equity context, "this measure captures non-price aspects of credit market conditions, such as debt covenants and quantity constraints." They find this measure to be strongly related to the amount of leverage used to finance private equity investments.

Turning to the empirical evidence on this channel, we first document a strong relation between private equity investment returns and the average innovation in market liquidity (as measured by Pastor-Stambaugh (JPE 2003)) during the investment's life. The average difference in performance for investments at the extreme deciles of market liquidity innovations is a striking 46% per year. This means that the 10% of the investments that had the best average liquidity conditions (compared to what was expected) during their life outperforms the 10% of the investments with the worst average liquidity conditions (compared to what was expected) during their life by 46% p.a. As there are other important determinants of private equity returns which may also be correlated with liquidity conditions, we verify and do confirm this simple result in a multiple regression setting, in which we control for investment characteristics and macroeconomic variables.

Figure 2: Annual performance by deciles of liquidity conditions
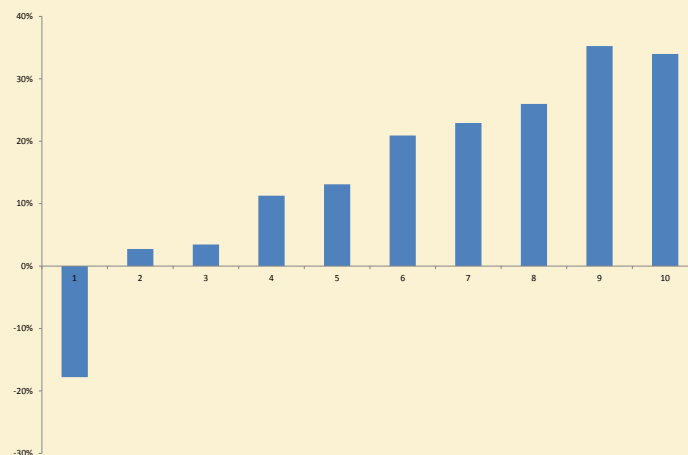
This is the average investment MIRR in each decile of the Pastor-Stambaugh (JPE 2003) liquidity condition variable.

Next, we test our conjecture that funding liquidity is the link between these two variables. We first show that returns are significantly related to the tightening in credit standards. A one-standard deviation increase in this measure of the deterioration in funding liquidity decreases the annual return by 16%. Second, when including both the measure of funding liquidity and that of market liquidity, we observe that funding liquidity absorbs half of the market liquidity effect. In addition, we conduct a time-series test using the aggregate cash flows of all the private equity investments each month. Consistent with the cross-sectional evidence, we find that net cash flows (dividends minus investments) are lower at times of tightening in credit standards and at times of worsening liquidity conditions.

Our results are important for two related reasons. First, they improve our understanding of the economic channel underlying the relationship between private equity returns and market liquidity. Market liquidity is found to be closely related to a measure of funding liquidity, which in turn is a determinant of the ease of refinancing for leveraged deals as shown by Axelson-Jenkinson--Stromberg-Weisbach (JF 2013). Second, these results provide empirical support for the theory of Brunnermeier-Pedersen (RFS 2009) relating funding liquidity to market liquidity. Our empirical evidence shows that there is indeed a negative relationship between a dry-up in funding liquidity (the tightening in credit standards) and innovations in market liquidity (the Pastor and Stambaugh measure).

Table 3: Explaining investment-annualized private equity MIRRs with liquidity conditions

P&S liquidity conditions	0.114***		0.051**	0.051
	(0.028)		(0.023)	(0.029)
Tightening of credit standards		–0.195***	–0.164***	–0.172***
		(0.031)	(0.026)	(0.035)
Industrial production growth				–0.003
				(0.041)
Delta credit spread				0.006
				(0.027)
Relative number of M&A deals				0.011
				(0.046)
Delta realized long term volatility				–0.022
				(–0.035)
Rm–Rf	0.068**	–0.013	–0.015	–0.019
	(0.032)	(0.031)	(0.029)	(0.048)
Growth investment	–0.036**	–0.042**	–0.041**	–0.040**
	(0.018)	(0.017)	(0.017)	(0.017)
Investment size	0.005	–0.001	0.000	0.001
	(0.014)	(0.009)	(0.000)	(0.010)
Fund size	–0.001	–0.003	–0.003	–0.002
	(0.004)	(0.004)	(0.005)	(0.004)
Country and industry fixed effects	yes	yes	yes	yes
Adj. R^2	0.093	0.118	0.123	0.124
N (1990–2007)	3,763	3,763	3,763	3,763

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

Details on the variables can be found in our original paper.

Raphael (1483–1520): Head of a Muse (about 11 inches). Italy, 1519. Private equity billionaire Leon Black bought this delicate beauty for $49.9 million dollars in 2009, *in the middle of the great recession*. Britain placed a temporary export ban on the painting, but finally allowed it leave the country in 2013. Does this high a price indicate revival of interest in acquiring Old Masters (or in Toryism or Laborism)? Are you in the right business? Who will pay this much for your masterpiece?

# Carry trades and global foreign exchange volatility

A miraculous free lunch for international investors seems to have been on offer over three decades: carry trades. In a carry trade an investor borrows money in currencies with low interest rates and invests this money in currencies with high interest rates. Following this most simple investment rule, excess returns of up to 10% p.a. could be gained over the last 30 years, without the input of any capital.

However, there is rarely a free lunch in financial markets. This is even less likely if markets are liquid, without barriers to trade (and any restrictions on short positions) and populated by sophisticated professionals. All this applies to the world's largest financial market, i.e. foreign exchange (FX). Accordingly, it does not seem very plausible that any limits to arbitrage should have hindered the elimination of the carry-trade's profitability. Therefore, if these profits are real and permanent, their rationale may be based on their inherent riskiness.

In Menkhoff-Sarno-Schmeling-Schrimpf (JF 2012), we indeed suggest that the returns to the carry trade can be understood as a compensation for risk. This means that the high carry-trade returns occur in "good times" for the investor, but that in "bad times" the carry trade performs poorly. As this idea has been investigated in various studies, we first present our specific argument, procedure and result before relating our research to other studies and showing some of its advantageous features.

Defining five carry-trade portfolios and market conditions

The examination of the carry trade follows a standard procedure: sort available currencies, in our case a maximum of 48, into portfolios of equal size, in our case five portfolios. Thus, investing in the Portfolio 5, i.e. currencies with relatively highest interest rates, and shorting the opposite Portfolio 1 results in the carry-trade portfolio. This allocation is updated at the end of each month. Repeating this exercise over 26 years—from November 1983 to August 2009—yields an excess return of more than 5% p.a. even after accounting for significant transaction costs (Figure 1). The outcome is only slightly worse when we reduce the sample to 15 developed countries, where FX markets display higher liquidity. Interestingly, these excess returns are neither related to risk factors of the CAPM and its variants nor to simple business cycle risk (as indicated by the shaded periods of NBER recessions in Figure 1 below).

Figure 1: Cumulative Carry-Trade Returns
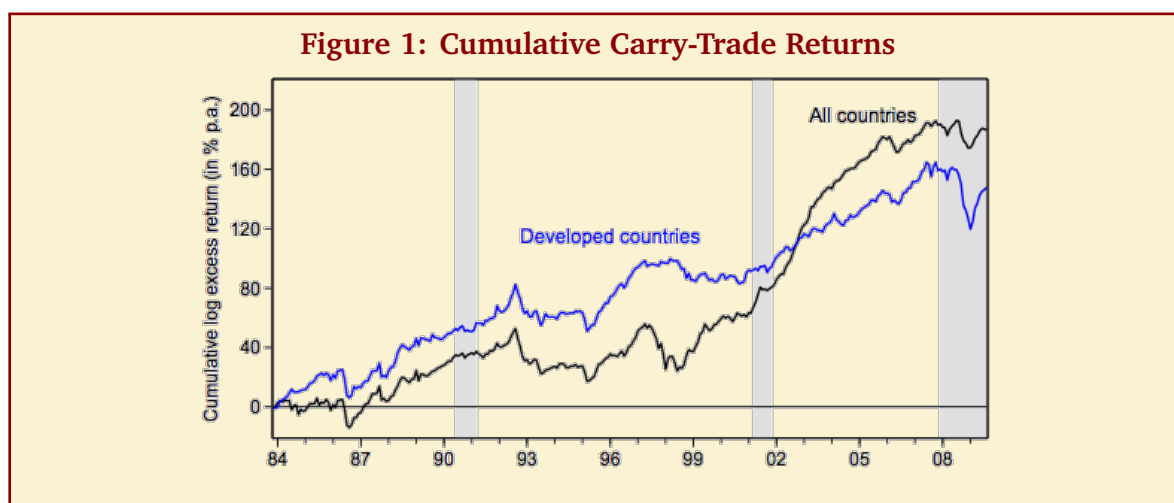
# Carry-trade returns are related to FX volatility

We introduce another measure of risk. Risk-averse investors may want to hedge themselves against unexpected changes (innovations) in market volatility. Investors demand currencies that hedge against this risk. We test whether the sensitivity of carry-trade returns to a measure of FX volatility can rationalize these returns in a standard asset pricing framework. For global FX volatility, we use the average of volatility innovations across all considered exchange rates, a measure which remains robust to some changes in its exact calculation.

Figure 2: Excess returns of carry trades depending on global FX volatility
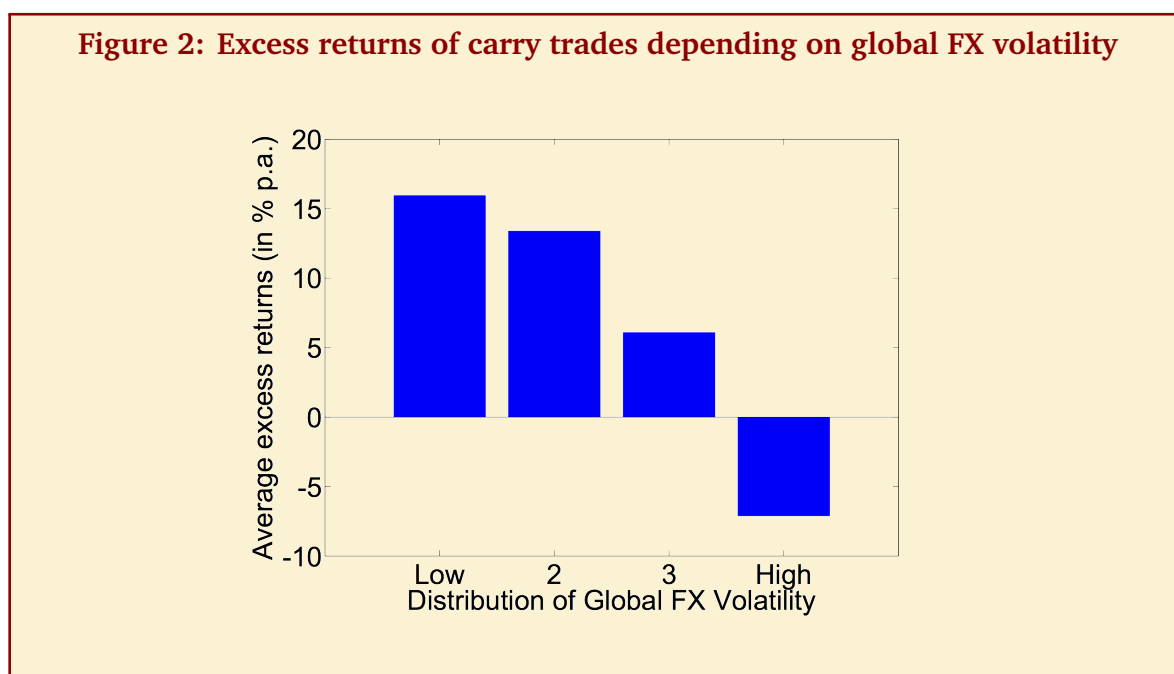
The usefulness of this risk measure is illustrated by Figure 2. Here all monthly results on global FX volatility are ordered into four groups, ranging from lowest volatility to highest volatility. The

average excess return of the carry-trade portfolios is positive except for the high volatility regime where returns become highly negative. This shows graphically the sensitivity of carry-trade returns to volatility risk.

In order to formally examine the pricing power of this risk factor we apply a standard linear asset pricing model. The graphical result of this analysis shows in Figure 3 that the realized mean returns of the five portfolios under consideration (where Portfolio 5 minus Portfolio 1 is the carry trade) nicely fit these portfolios' returns as expected from the model. In fact, the standard model relying on global FX volatility innovations as a risk factor explains about 95% of the portfolios' returns.

Figure 3: Portfolio excess returns: expected vs. realized
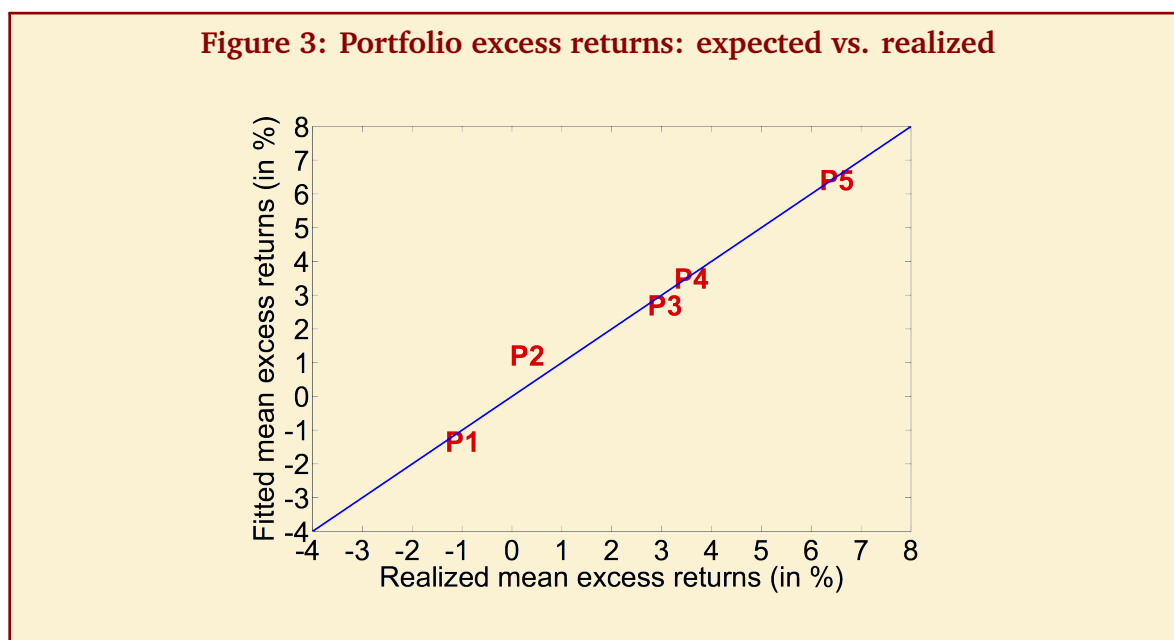
Liquidity risk is less important

We test the explanatory power of the global FX volatility risk factor in relation to other potential risk-based explanations. These other explanations refer to a lack of liquidity as risk, a specific carry-trade risk (HMLFX), skewness of carry-trade returns and Peso problems of carry trade strategies, i.e. the possibility of extremely rare and heavy losses.

A conventional measure of illiquidity in FX markets is the size of the bid-ask spread (BAS) and we calculate the BAS average across currencies, analogous to the procedure for global FX volatility. Second, in order to capture the crucial funding of carry-trade strategies Brunnermeier-Nagel--Pedersen (NBER 2009) suggest the TED spread as an illiquidity measure. The TED spread calculates the interest rate difference between 3 month interbank deposits and 3 month Treasury bills. Third, we take the liquidity measure introduced by Pastor and Stambaugh (P/S) for the US stock market as a proxy for FX market liquidity. We note that these three measures of (il)liquidity are clearly positive correlated to each other and to the global FX volatility, but the absolute values of correlation coefficients are always below 30% and hence imperfect.

When we replace in the asset pricing exercises volatility innovations by innovations of the three (il)liquidity measures we get reasonable results: the values of R^2 are (for the case of all countries) between 0.70 and 0.74, the factor prices have the expected sign, and are significantly different from zero for the BAS. However, these results are not quite as good as for the global FX volatility as risk factor. This impression is supported by asset pricing tests where we consider two risk factors, i.e. global FX volatility risk and (il)liquidity risk. We orthogonalize (il)liquidity risk with respect to volatility innovations, i.e. we consider only that part which is not explained volatility already. Table 1 shows the results based on risk price estimates (with standard errors in parentheses): the volatility risk factor carries a consistently negative risk price of about –0.06 to –0.08 and is highly significant, whereas the (il)liquidity risk factors are not significant when jointly included with volatility and a DOL factor (this dollar factor serves as a control variable).

Table 1: Explaining the cross-section of five carry trades

	DOL	BAS	TED	P/S	F/X VOL	R^2
Bid-Ask	0.21	0.01			–0.08**	0.98
	(0.31)	(0.02)			(0.04)	
TED Spread	0.21		–0.08		–0.06**	0.98
	(0.25)		(0.24)		(0.03)	
Pastor-Stambaugh (P/S)	0.18			–0.01	–0.08**	0.97
	(0.29)			(0.04)	(0.04)	

Standard errors are in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by one, two, and three stars, respectively.

These are the averaged stage-2 coefficients from enhanced Fama-Macbeth-style regressions, explaining five carry-trade excess returns with stage-1 exposures from 1983 to 2009. The independent variables are (currency exposures to) our FX volatility measure, plus a number of control measures: DOL, a dollar factor (a control); BAS, the prevailing average bid-ask spread across all currencies; and P/S, the Pastor-Stambaugh measure. For more detail, see Table VI in our original JF paper.

Volatility captures all relevant pricing information

Next we turn to a specific carry-trade risk factor (HMLFX) introduced by Lustig-Roussanov--Verdelhan (RFS 2011). This risk factor is the return to the carry-trade portfolio itself and is able to explain the pricing of the five portfolios introduced above. When comparing this risk factor with global FX volatility, we find that our volatility factor basically captures all of the information in HMLFX relevant for pricing the cross-section of carry portfolios. Transforming our non-tradable

volatility factor into a factor-mimicking portfolio for volatility innovations, we find that the resulting excess return is highly negatively correlated with HMLFX and even yields slightly lower pricing errors on the cross-section of carry portfolios. Moreover, the average return to the factor-mimicking portfolio of –1.3% p.a. is well in line with the average return to another volatility hedge portfolio: A zero-beta portfolio of long currency straddles (long position in both call and put options with the same strike price) that hedges against shocks to global volatility.

Skewness is unimportant

Skewness of returns is another risk factor suggested in the literature on carry trade (e.g., Brunnermeier-Nagel-Pedersen, NBER 2009). Even though there is some evidence in favor of skewness in our data, when assessing skewness as a systemic risk factor, results are not statistically significant. Similarly, we do not find clear evidence in favor of Peso problems as a source of carry-trade risk (Burnside-Eichenbaum-Kleshchelski-Rebelo (RFS 2011)). We also tested whether extreme volatility spikes drive our results. This was not the case, either.

Conclusion

In summary, our paper has proposed that a measure of innovations in global FX volatility as a systemic risk factor explaining carry-trade returns. This factor is a robust risk factor explaining the risk in carry-trade strategies well in comparisons to alternatives. Volatility risk works better than various forms of illiquidity risk or risk from skewness, and its power does not seem to be driven by neglected Peso problems. All of these explanations show that the high carry-trade returns are no free lunch, and among these explanations the risk of being exposed to innovations in global FX volatility is eminent.
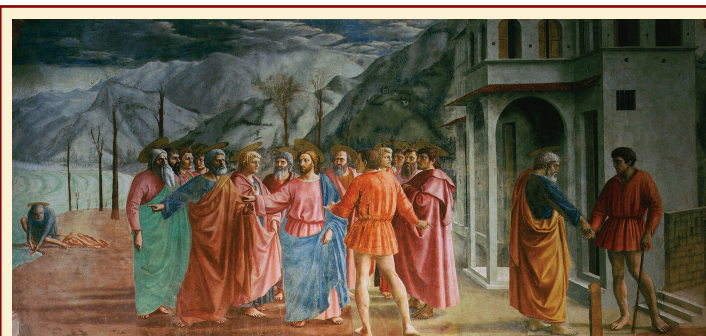
Masaccio (1401–1428): Tribute Money. Italy, 1427 Did we not already have this subject covered above? Were all the renaissance artists obsessed with the theme of Tribute Money? Again, here St. Peter takes direction from Jesus, produces the tax money from the mouth of a fish, and pays it to the tax collector to whom it rightfully belongs. In modern times, taxation is still full of miracles—just different types of miracles.

The "out-of-sample" performance of long-run risk models

The long-run risk model developed by Bansal-Yaron (JF 2004) has been a phenomenal success. One central feature of the model is that consumption and dividend growth contain a small long-run predictable component. A burgeoning literature shows that the model can explain various asset markets phenomena, including the equity premium puzzle, size and book-to-market effects, momentum, long-term return reversal in stock prices, risk premiums in bond markets, real exchange rate movements, and more. However, the evidence to date is mostly based on calibration exercises, in which researchers examine whether prices and returns generated by a calibrated model match actual prices and returns, or based on in-sample estimation, in which researchers choose model parameters to fit within the sample of asset returns. However, for an asset pricing model to be useful for practical applications, the model should be able to fit returns out-of-sample. The reason is that most practical applications are, in some sense, out of sample. For example, firms want to estimate cost of capital for future projects, portfolio and risk managers want to know the expected compensation for future risk, and academic researchers want to make risk adjustments to expected returns in future data. From this perspective, this paper provides an empirical analysis of the out-of-sample performance of the long-run risk models.

We study the "out-of-sample" performance of long-run risk models for explaining the equity premium puzzle, size and book-to-market effects, momentum, reversal, and bond returns of different maturity and credit quality. We examine both stationary and cointegrated versions of the models using annual data for 1931–2009. To evaluate out-of-sample fit, we use a traditional "rolling" estimation. We estimate the model parameters over an initial period and predict returns for the next year. We then roll the estimation period forward by one year and repeat the process.

The long-run models

Our stationary model specification follows Bansal-Yaron (JF 2004). In the model, consumption growth contains a small, highly persistent long-run risk component. The conditional volatility of consumption varies with time. There are three shocks: shocks to short-run consumption growth, shocks to long-run consumption growth, and shocks to consumption volatility.

Our cointegrated model follows Bansal-Dittmar-Lundblad (JF 2005) and Bansal-Dittmar-Kiku (RFS 2009). In the cointegrated model, the natural logarithms of aggregate consumption and dividend levels are cointegrated. This means that both consumption growth and dividend growth contain a highly persistent long-run risk component, but the two can't wander too far apart, so a weighted difference between them is stationary. The conditional volatility of consumption is time varying in this model as well. There are three shocks: shocks to the cointegrating relation between dividends and consumption, shocks to long-run consumption growth, and shocks to consumption volatility.

Empirical methods

For an asset pricing model to be useful in practical application, the model should be able to fit returns out-of-sample. We use the mean squared pricing error (MSE) as the criterion to evaluate models. A model that delivers the correct (conditional) mean of the future return minimizes the (conditional) MSE. We also compare the two components of the MSE—the expected pricing error and the error variance—of these different models. We compare the out-of-sample performance of long-run risk models with that of two simple models: a simple consumption beta model (CCAPM) and the classic Capital Asset Pricing Model of Sharpe (JF 1964). To evaluate the factors that drive the fit of the long-run risk models, we compare long-run risk models with models that suppress the consumption-related shocks. We also compare long-run risk models with models that restrict the risk premiums to identify structural parameters.

Main findings

We have the following main results. First, the cointegrated long-run risk model beats the original, stationary version in terms of out-of-sample performance. Thus, we argue that the out-of-sample performance suggests that cointegrated models should have a more prominent role in future research. Second, we find that the long-run risk models trump other models in explaining the Momentum Effect. At the same time, they perform poorly in explaining the relative returns to low-grade corporate and long-term government bonds. These results suggest that there are important missing factors in the simple, long-run risk models. Third, we find that both the short-run consumption risk factor and the long-run risk factor are important ingredients in the models' performance. Our evidence is that models that suppress the consumption-related shocks perform poorly (which indicates that the consumption shocks are important risk factors), and that the long-run risk models also perform better than the simple consumption-based model (CCAPM). Fourth, the mean squared errors of the long-run risk models are not substantially better than those

of the classical CAPM, except for Momentum. Fifth, when we restrict the risk premiums on the long-run risk models' factors to identify structural parameters of the model, we find that restricted model's overall performance is inferior to the classical CAPM, as the restriction increases the average out-of-sample pricing errors substantially, while sometimes reducing the pricing error variances.

Francisco Goya (1746–1828): Sleep of Reason Produces Monsters. Spain, 1799. Goya was the last and perhaps the coolest of the old masters. This famous etching was censored in its time, because it was considered a biting critique of 18th-century Spanish society, full of corruption and superstition. In this work, the sleeping "reason" is haunted by monsters and evil spirits. But Goya was not promoting reason alone. His caption reads "Imagination abandoned by reason produces impossible monsters; united with her, she is the mother of the arts and source of their wonders." Many of the other etchings in the series are both amusing and disturbing.