Sparse Signals in the Cross-Section of Returns*

Alex Chinco[†], Adam D. Clark-Joseph[‡], and Mao Ye[§]

February 16, 2017

Abstract

How do arbitrageurs find variables that predict returns? If a predictor lasts 30 days or more, then a clever arbitrageur can use his intuition to get the job done. But, what's an arbitrageur supposed to do if a predictor lasts 30 minutes or less? An arbitrageur's intuition is useless if the predictor decays before he can finish his morning coffee. Motivated by this observation, we show how arbitrageurs can find these sorts of rare, short-lived, "sparse" predictors by replacing intuition with a statistical procedure known as the LASSO. Using the LASSO boosts out-of-sample predictability in 1-minute returns by 23% relative to standard OLS-regression models. This out-of-sample predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some new factor. What's more, the predictors chosen by the LASSO correspond to real-world events: the lagged returns of stocks with announcements are 18.3% more likely to be used by the LASSO as predictors.

JEL Classification. C55, C58, G12, G14

Keywords. Return Predictability, Out-of-Sample Fit, Sparsity, The LASSO

---

*We have received many helpful comments and suggestions from John Campbell, Victor DeMiguel, Xavier Gabaix, Andrew Karolyi, Bryan Kelly, Maureen O'Hara, Vassilis Papavassiliou, Ioanid Rosu, Thomas Ruchti, Gideon Saar, Heather Tookes, Sunil Wahal, and Brian Weller as well as seminar participants at the University of Illinois Urbana-Champaign, the 11th Annual Central Bank Conference on the Microstructure of Financial Markets, and the 2016 AFA Annual Meetings. This research is supported by National Science Foundation grant 1352936 (joint with the Office of Financial Research at the U.S. Treasury Department). This work also uses the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. We thank David O'Neal of the Pittsburgh Supercomputer Center for his assistance with supercomputing, which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program.

Current Version: http://www.alexchinco.com/sparse-signals-in-cross-section.pdf

[†]University of Illinois at Urbana-Champaign, alexchinco@gmail.com, (916) 709-9934.

[‡]University of Illinois at Urbana-Champaign, adcj@illinois.edu, (217) 244-1536.

[§]University of Illinois at Urbana-Champaign, maoye@illinois.edu, (217) 244-0474.

1 Introduction

Financial economists have been looking for variables that predict future stock returns for as long as there have been financial economists. For example, Banz (1981) shows that a company's size predicts its future returns, Jegadeesh and Titman (1993) show that a stock's past returns predict its future returns, and Cohen and Frazzini (2008) show that the earnings of a company's major customers predict the company's future returns. Finding these predictors actually involves two distinct steps: identification and estimation. First, researchers have to identify a potential predictor, x_t. Then, they have to estimate its quality,

$$r_{n,t+1} = \hat{\theta}_0 + \hat{\theta}_1 \cdot x_t + \epsilon_{n,t+1}, \tag{1}$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are estimated coefficients and $r_{n,t+1}$ is the return on the nth stock. If x_t is a good predictor, then $\hat{\theta}_1$ will be statistically significant and the associated R^2 will be large.

In the past researchers have always relied on their intuition when trying to identify the predictors to include in their regressions, but modern financial markets are big, fast, and complex. Predictability does not always occur at scales that are easy for researchers to intuit, making the standard approach to identifying predictors problematic. For instance, the lagged returns of Family Dollar Corp. were a significant predictor for the returns of more than 20% of all oil and gas stocks during a 20-minute stretch on October 6th, 2010. Can a researcher really fish this particular variable out of a sea of spurious predictors using intuition alone? And, how exactly is he supposed reel in this variable in under 19 minutes?

This paper uses the LASSO, rather than intuition, to identify rare, short-lived, "sparse" predictors like Family Dollar in the example above. Estimating the LASSO, which stands for the least absolute shrinkage and selection operator, is just like estimating the ordinary-least-squares (OLS) regression in Equation (1) except for one key difference: there is an additional penalty term that sets what would have been small OLS coefficients to be exactly zero. This penalty term prunes away the least significant right-hand-side variables to reveal the handful of relevant predictors. Because of this penalty term, the LASSO is well defined even when there are many more right-hand-side variables than sample periods.

We show that using the LASSO to identify sparse signals boosts out-of-sample predictability in one-minute returns by 23% relative to standard OLS regressions.

And, we verify that this out-of-sample predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some new, persistent factor. When we examine the economic origins of the LASSO's success, we find that it comes from identifying predictors that subsequently realize news announcements. A stock with a news announcement in minute t is 18.3% more likely to be used by the LASSO as a predictor in minutes $\{(t-30), \ldots, (t-1)\}$. This link between the LASSO's predictors and future news announcements provides an economic foundation for applying machine-learning techniques, like the LASSO, to empirical finance.

Estimation Strategy. We begin in Section 2 by describing how the LASSO works. We want to forecast NYSE-listed stock's returns each minute using the lagged returns of other NYSE stocks. So, why not just add in the lagged returns of other NYSE stocks when estimating the predictive regression in Equation (1)? There are 2,191 NYSE stocks in our data for October 2010, so the resulting estimation problem would have 2,192 free parameters: one for the intercept and one for the lagged return of each NYSE stock in the previous minute. Estimating all of these free parameters using a standard OLS regression would require at least 2,192 minutes of data, which is nearly 6 trading days! Most cross-stock signals would be gone before we could generate a well-defined OLS estimate. To shorten the required sample length, we need to take a different approach and focus on only the most important cross-stock predictors. The LASSO allows us to do just that.

Using the LASSO means solving the optimization problem below,

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta} \in \mathbf{R}^{2,192}} \left\{ \frac{1}{2 \cdot T} \cdot \sum_{t=1}^{T} \left(r_{n,t+1} - \vartheta_0 - \sum_{n'=1}^{2,191} \vartheta_{n'} \cdot r_{n',t} \right)^2 + \lambda \cdot \sum_{n'=1}^{2,191} |\vartheta_{n'}| \right\}, \tag{2}$$

where $r_{n,t}$ is the nth stock's return in minute t, $\hat{\boldsymbol{\vartheta}}$ is a $(2{,}192 \times 1)$-dimensional vector of estimated coefficients, T is the number of minutes in the sample, and λ is a penalty parameter. In all of our analysis, we use θ to denote OLS coefficients and ϑ to denote LASSO coefficients. Equation (2) looks complicated at first, but it is not. It is a simple extension of an OLS regression. In fact, if you ignore the right-most term—the penalty function, $\lambda \cdot \sum_{n'} |\vartheta_{n'}|$—then this optimization problem would simply *be* an OLS regression. But, it is this penalty function that is the secret to the LASSO's success, allowing the estimator to give preferential treatment to the largest coefficients and completely ignore the smaller ones.

To see how, consider the solution to Equation (2) when the right-hand-side variables are uncorrelated and have unit variance:

$$\hat{\vartheta}_{n'} = \mathrm{sgn}[\hat{\theta}_{n'}] \cdot (|\hat{\theta}_{n'}| - \lambda)_{+}.$$

Here, $\hat{\theta}_{n'}$ represents what the standard OLS coefficient would have been if we had an infinite amount of data, $\mathrm{sgn}[x] = {}^{x}/_{|x|}$, and $(x)_{+} = \max\{0, x\}$. On one hand, if OLS would have estimated a large coefficient, $|\hat{\theta}_{n'}| \gg \lambda$, then the LASSO is going to deliver a similar estimate, $\hat{\vartheta}_{n'} \approx \hat{\theta}_{n'}$. On the other hand, if OLS would have estimated a sufficiently small coefficient, $|\hat{\theta}_{n'}| < \lambda$, then the LASSO is going to pick $\hat{\vartheta}_{n'} = 0$. Thus, because the LASSO can set all but a handful of coefficients to zero, it can be used to identify the most important predictors even when the sample length is much shorter than the number of right-hand-side variables, $T \ll 2{,}192$. Morally speaking, if only $K \ll 2{,}192$ of the predictors are non-zero, then you should only need a few more than K observations to identify these few important coefficients.

Out-of-Sample Predictability. Next, in Section 3 we show that the LASSO boosts out-of-sample predictive power. As a benchmark, we start by estimating rolling autoregressions with 30 minutes of data and find that the average out-of-sample adjusted R^2 is 8.17% for NYSE stocks. That is, on average 8.17% of the total variation in an NYSE stock's minute-by-minute returns can be accounted for by studying that stock's past returns, and only that stock's past returns. We then consider the predictive power that comes from including other stocks' lagged returns over the previous 3 minutes via the LASSO. This means using 30 minutes of data to both identify the handful of significant predictors from among $1 + (3 \times 2{,}191) = 6{,}574$ possibilities in October 2010. Including the LASSO's return forecast boosts the out-of-sample adjusted R^2 by 23%, from an adjusted $R^2 = 8.17\%$ to an adjusted $R^2 = 10.05\%$.

We verify that the LASSO's out-of-sample predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some new, persistent factor. On average, the LASSO only selects around 11 predictors to make each stock's return forecast for a given minute. This is only around 0.5% of the roughly 2,000 stocks that it can choose from. What's more, the set of significant predictors changes rapidly. If the LASSO is using the lagged returns of the Family Dollar to predict Exxon's future returns right now, then there is less than a 10% chance that the LASSO will still be using Family Dollar's lagged returns 5 minutes later.

Economic Origins. Finally, in Section 4, we examine the economic origins of the LASSO's out-of-sample predictive power. We document that, even though we fit the LASSO separately when making return forecasts for different stocks, the LASSO still tends to identify the same predictors at the same time when making these separate forecasts. So, for example, if the LASSO is using the lagged returns of Family Dollar to forecast Exxon's returns, then it is also much more likely to be using Family Dollar when making return forecasts for other stocks, such as British Petroleum, Chevron, or Motorola. And, we show that this predictor clustering is related to subsequent news announcements. A stock with a news announcement in minute t is 18.3% more likely to be used by the LASSO as a predictor in minutes $\{(t-30), \ldots, (t-1)\}$. In other words, the LASSO does not proxy for news announcements, it predicts them. Taken together, these results suggest that traders can use a penalized regressions, like the LASSO, to identify information about real-world events that is missed by standard regression techniques, like OLS.

1.1 Related Literature

The paper builds on several strands of the statistics and asset-pricing literatures.

Return Predictability. First, our paper builds on a large literature exploring stock-return predictability. Researchers have found negative autocorrelation between past and future returns at horizons shorter than 3 months (Fama, 1965; Lo and MacKinlay, 1990; Conrad et al., 1991), positive autocorrelation at horizons between 3 and 12 months (Jegadeesh, 1990; Jegadeesh and Titman, 1993; Asness, 1994; Chan et al., 1996; Carhart, 1997), and negative autocorrelation at horizons longer than a year (De Bondt and Thaler, 1985). Moskowitz et al. (2012) show that momentum is both a time-series as well as a cross-sectional phenomenon. Lo and MacKinlay (1990) and Boudoukh et al. (1994) document that there is a lead-lag relationship in price movements across stocks. Numerous behavioral (Barberis et al., 1998; Daniel et al., 1998; Hong and Stein, 1999) and rational (Berk et al., 1999; Johnson, 2002; Ahn et al., 2003; Liu and Zhang, 2008) explanations for this return predictability have been suggested. Our analysis suggests that the source of this predictability can change over time. So, the explanation for the return predictability observed in the data may not involve a set of persistent factors.

Out-of-Sample Fit. Our paper also builds on a related line of work that explores the out-of-sample performance of common predictors. For example, Campbell and Thompson (2008) show that many common predictors only add out-of-sample predictive power with additional restrictions. DeMiguel et al. (2009) and Zhu and Zhou (2009) document that, when there is uncertainty about the true asset-pricing model, naïve trading strategies and technical trading rules can outperform model-based predictors out-of-sample. And, DeMiguel et al. (2014) show that simple vector-autoregressive (VAR) models can boost out-of-sample return predictability.

A broad theme in this literature is that simple predictors are often better at forecasting out-of-sample returns when there is uncertainty about the underlying asset-pricing model. A precise, fine-tuned predictor can do a very bad job if it has been fine-tuned for precisely the wrong asset-pricing model. One way to think about how the LASSO is adding out-of-sample predictive power is by avoiding this predictor-model mismatch by silencing unhelpful predictors with the penalty term.

Sparsity. Our paper relates to new work on the role of sparsity in economics and finance. For example, Gabaix (2014) proposes that boundedly rational traders view the world through sparse mental models. Traders in this framework prune away the least important predictors by imposing an ℓ_1 penalty. Using the LASSO to identify sparse signals in the cross-section of returns precisely corresponds to using this same ℓ_1 pruning rule. But, while in Gabaix (2014) this pruning rule emerges from traders' bounded rationality, in the current paper this pruning rule is a way to identify sparse signals that are actually in the data. From the opposite perspective, Chinco (2015) shows that, if traders have to uncover sparse signals in past market data, then there are information-theoretic limits to how quickly they can interpret what the market is telling them. In a closely related paper, Manela and Moreira (2015) use a support-vector regression to predict aggregate stock-market returns at much lower frequency using the text of front-page news stories.

The LASSO. Finally, there is a large statistics literature examining the performance of penalized regressions, like the LASSO. The LASSO was introduced in Tibshirani (1996). Hastie et al. (2001) provide a general introduction to the LASSO. Meinshausen and Yu (2009) give an excellent overview of how well these LASSO-type estimators extend to settings with correlated right-hand-side variables. Other regressions with ℓ_1 penalties exist in the literature (Efron et al., 2004; Zou and Hastie, 2005; Candes and Tao, 2007). Belloni et al. (2012, 2014) and Chernozhukov et al. (2015)

apply the LASSO to estimate treatment effects in settings where there are a large number of (potentially weak) instruments. More generally, DeMiguel et al. (2009) show that the penalized-regression framework nests a large number of commonly used estimators in empirical finance.

2 Estimation Strategy

This section describes how we use the LASSO to identify sparse signals in the cross-section of returns. We start by outlining an OLS model, which captures the benchmark level of return predictability if we do not consider sparse signals. Then, we discuss the LASSO and how it differs from the benchmark OLS model. Finally, we highlight the necessary conditions for the LASSO to add out-of-sample predictive power when forecasting stock returns.

2.1 OLS Regression

Our main finding is that, by identifying sparse signals in the cross-section of returns, the LASSO boosts out-of-sample predictive power by 23%. In order to make this claim, we first need to know how predictable minute-by-minute stock returns would be if we did not consider any sparse signals. We estimate this benchmark level of predictability using an OLS regression.

Fit and Prediction. As a benchmark model, we make a return forecast in minute $(t+1)$ for each NYSE-listed stock by fitting an autoregressive model with 3 lags—that is, an AR(3) model—to the stock's returns over the previous 30 minutes,

$$r_{n,\tau} = \hat{\theta}_0 + \textstyle\sum_{\ell=1}^{3} \hat{\theta}_\ell \cdot r_{n,\tau-\ell} + \epsilon_{n,\tau} \quad \text{for } \tau \in \{\, t-29, t-28, \, \ldots, t \,\},$$

where $\hat{\theta}_0$ and $\{\hat{\theta}_\ell\}$ are estimated coefficients, $r_{n,t}$ denotes stock the nth stock's return in minute t, and $\epsilon_{n,t}$ is the regression residual. Figure 1 outlines the timing of the OLS regressions in our baseline specification. For example, if we wanted to make a return forecast for IBM at 12:00pm, then we would fit an AR(3) model to IBM's returns over the previous 30 minutes, $t \in \{\, \text{11:30am}, \text{11:31am}, \, \ldots, \text{11:59am} \,\}$.

To predict the stock's returns in minute $(t+1)$, we use the fitted coefficient values and the final 3 minutes of data, $\{\, t-2, t-1, t \,\}$,

$$f^{\text{OLS}}_{n,t} \equiv \mathrm{E}_t[r_{n,t+1}] = \hat{\theta}_0 + \textstyle\sum_{\ell=1}^{3} \hat{\theta}_\ell \cdot r_{n,t-\ell+1}. \tag{3}$$

Timing of OLS Regressions

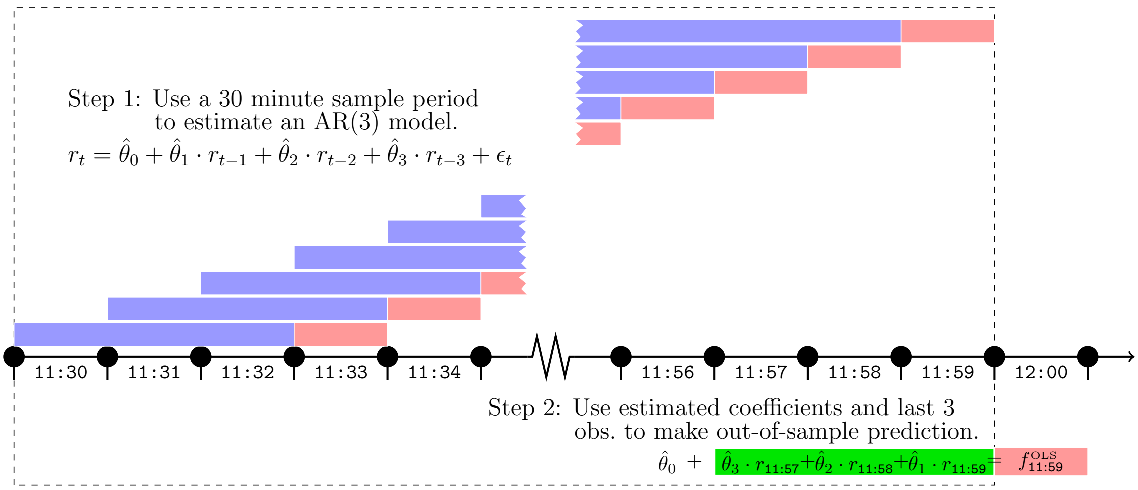

Figure 1: *To make an out-of-sample prediction using the baseline OLS model in minute* $(t+1) =$ *12:00pm, we first fit an autoregressive model with* 3 *lags using the stock's returns in the previous* 30 *minutes. Then, we use the fitted coefficients to predict the stock's return in minute* $(t+1) =$ *12:00pm, referring to the prediction as* $\mathrm{E}_t[r_{t+1}] \equiv f_t^{\text{OLS}}$.

Continuing with the example from above, to make IBM's return forecast for 12:00pm, we use the fitted coefficients from the previous 30 minutes and IBM's returns in minutes 11:57am, 11:58am, and 11:59am. Let $\tilde{\mu}_n^{\text{OLS}}$ and $\tilde{\sigma}_n^{\text{OLS}}$ represent the mean and standard deviation of this out-of-sample prediction for the nth stock over the course of the entire sample period.

Prediction Accuracy. To test whether these out-of-sample predictions are good or bad, we regress the realized returns in the 31st minute on the normalized return forecast for each stock,

$$r_{n,t+1} = \tilde{a}_n + \tilde{b}_n \cdot \left(\tfrac{f_{n,t}^{\text{OLS}} - \tilde{\mu}_n^{\text{OLS}}}{\tilde{\sigma}_n^{\text{OLS}}} \right) + e_{n,t+1}, \tag{4}$$

where $\tilde{a}_n$ and $\tilde{b}_n$ are estimated coefficients, $r_{n,t+1}$ denotes stock n's realized return in minute $(t+1)$, $f_{n,t}^{\text{OLS}}$ denotes our prediction at time t of stock n's return in minute $(t+1)$ using an autoregressive model, and $e_{n,t+1}$ is the regression residual. If $|\tilde{b}_n|$ or the R^2 associated with this regression in Equation (4) is large, then the OLS model does a good job of forecasting the next minute's return for stock n. To be clear, this means running separate regressions for each stock. So, for example, in October 2010 we run 2,191 of these regressions—one for each of the 2,191 NYSE-listed stocks in our sample in October 2010.

Since we are running a separate regression for each stock, we normalize the out-of-sample predictions by their stock-specific standard deviation to make the coefficients easier to compare. We can interpret the resulting slope coefficient, $\tilde{b}_n$, as the average return per minute to a time-series momentum strategy à la Moskowitz et al. (2012):

$$\tilde{b}_n = \tfrac{1}{T \cdot \tilde{\sigma}_n^{\text{OLS}}} \cdot \textstyle\sum_{t=1}^{T} (f_{n,t}^{\text{OLS}} - \tilde{\mu}_n^{\text{OLS}}) \cdot r_{n,t+1}. \tag{5}$$

Of course, this estimate ignores trading costs, which are (to put it mildly) substantial when rebalancing once every minute rather than once every month like in the original paper. On top of this, the market-timing strategy depends on knowing the distribution of the OLS model's out-of-sample prediction for each stock, $\tilde{\mu}_n^{\text{OLS}}$ and $\tilde{\sigma}_n^{\text{OLS}}$, even though this information is not known at the beginning of the sample period. In Appendix B, we examine whether it is possible to profitably trade on this signal.

Alternative Specifications. In our baseline specification, we include only a stock's lagged returns in minutes t, $(t-1)$, and $(t-2)$ when trying forecast that stock's returns in minute $(t+1)$. But, there are other variables that might help forecast a stock's future returns. For instance, think about the return on the market portfolio. It is easy to add the lagged values of such control variables into our baseline specification,

$$r_{n,\tau} = \hat{\theta}_0 + \textstyle\sum_{\ell=1}^{3} \hat{\theta}_\ell \cdot r_{n,\tau-\ell} + \sum_{\ell=1}^{3} \hat{\boldsymbol{\gamma}}_\ell^\top \cdot \mathbf{x}_{n,\tau-\ell} + \epsilon_{n,\tau} \quad \text{for } \tau \in \{\, t-29, \ldots, t\}, \tag{6}$$

where $\mathbf{x}_{n,t-\ell}$ is a vector of additional control variables and $\hat{\boldsymbol{\gamma}}_\ell$ is the associated vector of coefficients.

In addition to considering the return on the market portfolio, we also include the strongest pairwise predictor for stock n yesterday as a control variable when forecasting stock n's returns today. We are interested in whether or not the LASSO boosts out-of-sample forecasting power by quickly identifying sparse predictors in the cross-section of returns. So, to show that the adverb "quickly" is really warranted here, it is useful to include in the benchmark model the returns of the most obvious predictor in a longer sample period. If the LASSO increases out-of-sample predictive power relative to this model, then it must be identifying a new predictor with a shorter time horizon.

2.2 The LASSO

With this benchmark in hand, we next consider the impact of using other stocks' lagged returns over the previous 3 minutes when predicting the nth stock's return in minute $(t+1)$ via the LASSO. There are roughly $N \approx 2{,}000$ NYSE-listed stocks in our sample each October. So, using the LASSO means using 30 minutes of data to both identify and estimate the few significant predictors from among $1 + (N \times 3) \approx 6{,}000$ possibilities each month, a task that would clearly be impossible using OLS.

Fit and Prediction. For each of the NYSE-listed stocks in our sample each October, we compute a series of LASSO estimates using the same rolling 30-minute windows as we did for the OLS-based approach. The LASSO solves the optimization problem below,

$$\hat{\boldsymbol{\vartheta}} = \operatorname*{arg\,min}_{\boldsymbol{\vartheta} \in \mathbf{R}^{1+3\cdot N}} \left\{ \frac{1}{2 \cdot 30} \cdot \sum_{\tau=t-29}^{t} \left(r_{n,\tau} - \vartheta_0 - \sum_{n'=1}^{N} \sum_{\ell=1}^{3} \vartheta_{n',\ell} \cdot r_{n',\tau-\ell} \right)^2 + \lambda \cdot \sum_{n'=1}^{N} \sum_{\ell=1}^{3} |\vartheta_{n',\ell}| \right\},$$

where $\vartheta_{n,0}$ and $\{\vartheta_{n',\ell}\}$ are estimated coefficients, $r_{n,t-\ell}$ denotes stock the nth stock's return ℓ minutes ago, λ is a penalty parameter, and N is the number of stocks in a given month. To see whether the LASSO's coefficient estimates contain useful information, we create an out-of-sample prediction for each 30-minute training sample just as before,

$$f_{n,t}^{\text{LASSO}} \equiv \mathrm{E}_t[r_{n,t+1}] = \hat{\vartheta}_0 + \textstyle\sum_{n'=1}^{N} \sum_{\ell=1}^{3} \hat{\vartheta}_{n',\ell} \cdot r_{n',t-\ell+1}. \tag{7}$$

And, just as before, let $\tilde{\mu}_n^{\text{LASSO}}$ and $\tilde{\sigma}_n^{\text{LASSO}}$ represent the mean and standard deviation of this out-of-sample prediction for the nth stock.

Recall that the solution to the LASSO's optimization problem when the right-hand-side variables are uncorrelated and have unit variance is given by

$$\hat{\vartheta}_{n',\ell} = \mathrm{sgn}[\hat{\theta}_{n',\ell}] \cdot (|\hat{\theta}_{n',\ell}| - \lambda)_+,$$

where $\hat{\theta}_{n',\ell}$ represents what the OLS coefficient would have been given enough data, $\mathrm{sgn}[x] = {}^{x}\!/\!_{|x|}$, and $(x)_+ = \max\{0, x\}$. If OLS would have estimated a large coefficient, $|\hat{\theta}_{n',\ell}| \gg \lambda$, then the LASSO will deliver a similar estimate, $\hat{\vartheta}_{n',\ell} \approx \hat{\theta}_{n',\ell}$. When you look all the way to the right or to the left in Figure 2, you see that the solid line denoting the LASSO estimate and the dotted line denoting the OLS estimate are close. By contrast, if OLS would have estimated a sufficiently small coefficient, $|\hat{\theta}_{n',\ell}| < \lambda$, then the LASSO will pick $\hat{\vartheta}_{n',\ell} = 0$. This corresponds to the flat region in

Relationship Between LASSO and OLS Estimates

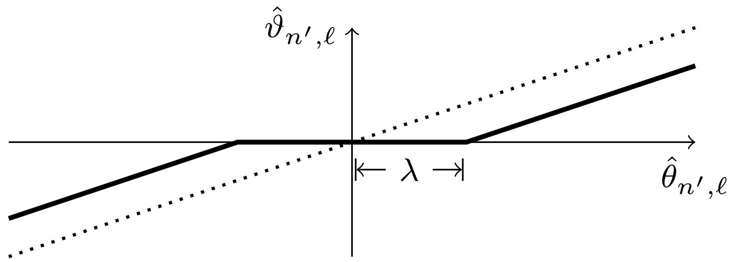

Figure 2: *x-axis: OLS-regression coefficient in an infinite sample. y-axis: Penalized-regression coefficient from the LASSO. Dotted: $x = y$ line. Reads: "If an OLS regression would have estimated a small coefficient value given enough data, $|\hat{\theta}_{n',\ell}| < \lambda$, then the LASSO will set $\hat{\vartheta}_{n',\ell} = 0$."*

Figure 2. Note that the LASSO could select stock n's lagged returns when trying to forecast stock n's future returns. In principle the LASSO could identify precisely the same predictors as we use in the some of the OLS regressions from Subsection 2.1.

A physical analogy gives some intuition about how the LASSO works. The LASSO identifies the most important predictors using a penalty function in the same way that squinting helps you make out the shape of distant objects. Suppose that you are standing on the beach and trying to figure out whether a dark spot on the horizon is a sailing ship or a cloud. The natural reaction in this situation is to squint. But, why is this so? By squinting and slightly closing your eyes, you actually make it harder for yourself to see colors and quick movements. In essence, squinting penalizes these weak signals that cannot tell you much about whether the mysterious blob on the horizon is a ship or a cloud. Both objects are going to be dark and slow moving at this distance. When you squint, the only things left for you to see are the most relevant details about the shape of the object. Is it triangular? Does it have a mast? Thus, squinting penalizes the least important signals, making it possible to identify the nature of the object on the horizon at a greater distance.

Prediction Accuracy. To test whether or not the LASSO adds out-of-sample predictive power relative to the OLS model, we regress each stock's realized returns in minute $(t+1)$ on both the OLS model's normalized return forecast and the LASSO's normalized return forecast for minute $(t+1)$,

$$r_{n,t+1} = \tilde{a}_n + \tilde{b}_n \cdot \left(\tfrac{f_{n,t}^{\text{OLS}} - \tilde{\mu}_n^{\text{OLS}}}{\tilde{\sigma}_n^{\text{OLS}}}\right) + \tilde{c}_n \cdot \left(\tfrac{f_{n,t}^{\text{LASSO}} - \tilde{\mu}_n^{\text{LASSO}}}{\tilde{\sigma}_n^{\text{LASSO}}}\right) + e_{n,t+1}, \tag{8}$$

where $\tilde{a}_n$, $\tilde{b}_n$, and $\tilde{c}_n$ are estimated coefficients, $r_{n,t+1}$ denotes the nth stock's realized return in minute $(t+1)$, $f_{n,t}^{\text{OLS}}$ and $f_{n,t}^{\text{LASSO}}$ denote our predictions of the nth stock's return in minute $(t+1)$ using the OLS model and the LASSO respectively, and $e_{n,t+1}$ is the regression residual. If $|\tilde{c}_n|$ is large or the adjusted R^2 associated with this regression in Equation (8) is large relative to the adjusted R^2 from the regression in Equation (4), then the LASSO adds out-of-sample predictive power for stock n. Again, to estimate these coefficients, we run separate regressions for each stock. In our main analysis, we measure the fit of each model by comparing the resulting adjusted R^2s from these regressions. Since these are nested models, we can also use an F-test to verify that the LASSO's return forecast is increasing out-of-sample predictive power.

2.3 Betting on Sparsity

We conclude this section by highlighting the situations where the LASSO's return forecast adds out-of-sample predictive power relative to the benchmark OLS model. The easiest way to do this is by discussing a candidate data-generating process and showing what features it must have for the LASSO to be successful. Here, we simply discuss the data-generating process. In Appendix A, we analyze this statistical model in more detail via simulations.

Data-Generating Process. Consider a market with $N = 100$ stocks and $T = 1{,}150$ trading periods. In each trading period, the returns of all $N = 100$ stocks are governed by the lagged returns of a subset of K stocks, $\mathcal{K}_t$, together with an idiosyncratic shock,

$$r_{n,t} = \vartheta \cdot \textstyle\sum_{n' \in \mathcal{K}_t} r_{n',t-1} + 0.001 \cdot \epsilon_{n,t},$$

where $\vartheta \geq 0$ is a scalar denoting the magnitude of the cross-stock predictability and $\epsilon_{n,t} \overset{\text{iid}}{\sim} \text{N}(0,1)$. We consider single-period lags for simplicity, but the results easily generalize to more complicated lag structures. The collection of K stocks changes over time, leading to the time subscript on $\mathcal{K}_t$. We assume that a signal changes in a given period with probability δ, so each signal lasts ${}^{(1-\delta)}/_{\delta}$ trading periods on average.

Winning Bet. Roughly speaking, the LASSO will add out-of-sample return predictability relative to an autoregressive benchmark if the following three conditions are met: 1) $\vartheta > 0.001$; 2) $K < 30$; and 3) δ is relatively large. If $\vartheta \leq 0.001$, then any sparse signals would be drowned out by the idiosyncratic noise. If there are only a few (that is, $K < 30$) important predictors with coefficients $\vartheta \gg 0.001$ in any 30-minute time window, then the LASSO will be able to identify and estimate these sparse signals, providing useful information when trying to forecast returns. In the simulations in Appendix A, we assume that $\vartheta = 0.15$ and that there is a 1% chance that each signal changes every period, so each signal lasts $^{(1-0.01)}/_{0.01} = 99$ trading periods on average.

Losing Bets. But, if any of these three conditions fails to hold, then the LASSO will not add out-of-sample predictive power. It is possible, in other words, to bet on sparsity and lose. First, if ϑ is too small, then there will not be any cross-stock signals for the LASSO to estimate. Hence, estimating the LASSO will not add any out-of-sample forecasting power. Second, if $K > 30$, then there would be too many cross-stock signals to estimate using only 30 data points. As a result, the LASSO would not add any out-of-sample forecasting power. For instance, suppose that the only relevant forecasting variable was the return on the market portfolio, which would mean that $\vartheta = {}^1/_N$ for N large. If this were the case, then the LASSO would not add any out-of-sample forecasting power.

Last but not least, if δ is not very large and the correlation structure is quite persistent, then the LASSO's return forecast will not contain any additional information that is not already in the OLS model's return forecast. Budish et al. (2015) show that the correlation structure of stock returns is different at different horizons. Using the LASSO at the one-minute horizon is doing more than just identifying a different short-term correlation structure. This different structure is already captured by the OLS model's return forecast. In order for the LASSO's return forecast to boost the accuracy of our out-of-sample predictions, it has to be the case that the choice of relevant predictors, $\mathcal{K}_t$, is not stable. It has to be the case that the set of relevant predictors, $\mathcal{K}_t$, is constantly changing; the lagged returns of Family Dollar are helpful for predicting Exxon's future returns right now but they will not be in 2 hours.

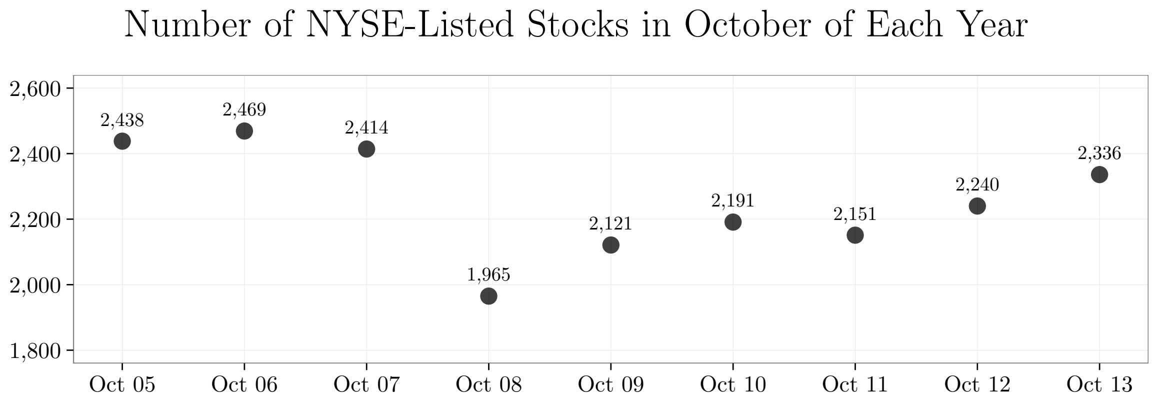

Figure 3: *Number of NYSE-listed stocks in the TAQ database during October of each year which had prices exceeding* \$5 *at the start of the month and which were traded every day of the month.*

3 Out-of-Sample Predictability

We find that the minute-by-minute returns of NYSE-listed stocks are 23% more predictable out-of-sample when using the LASSO. This section describes how we estimate this number and gives evidence that this out-of-sample predictive power comes from quickly identifying the right predictors at the right time.

3.1 Data Description

Our data consist of minute-level returns of NYSE-listed stocks collected from the TAQ database for the months of October in each year from 2005 to 2013. We restrict the sample to stocks which had a price exceeding \$5 at the close of the last trading day in September and which were traded every day in October. Figure 3 shows the number of observations in our sample each year. The first 30-minute window we consider each day is $t \in \{\,$10:36am, $\ldots$, 11:05am$\,\}$ and the last window is $t \in \{\,$3:29pm, $\ldots$, 3:58pm$\,\}$. This means that our first out-of-sample prediction each day is for minute 11:06am and our last out-of-sample prediction each day is for minute 3:59pm. We use the first 45 minutes of the trading day to fit the LASSO's penalty parameter, λ. We drop the last minute of the trading day due to avoid distortions due to the NYSE closing auction.

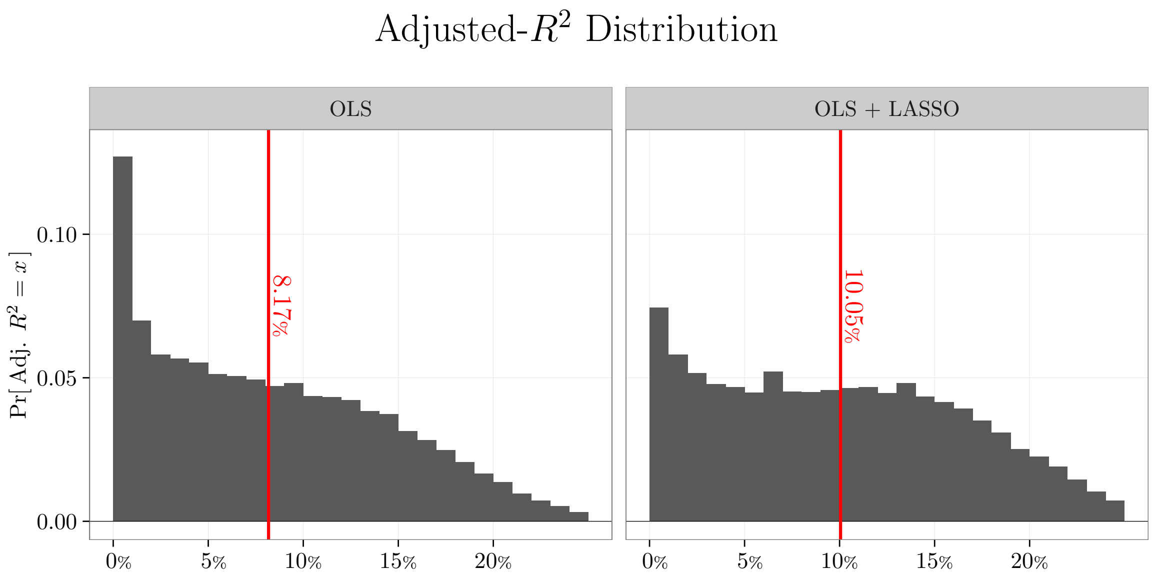

Figure 4: *Distribution of adjusted R^2s from the forecasting regressions in Equations* (4) *and* (8)*. Black bars: Probability that the adjusted R^2 from a single out-of-sample forecasting regression falls within a* 1%*-point interval. Red vertical line: Average adjusted R^2 from these regressions corresponding to the point estimates in the bottom row of Table 1. Left panel: Out-of-sample prediction made using OLS as in Equation* (4)*. Right panel: Out-of-sample predictions made using both OLS and the LASSO as in Equation* (8)*. Reads: "Including the LASSO's return forecast increases out-of-sample predictive power by* $^{10.05}/_{8.17} - 1 = 23\%$ *relative to the benchmark OLS model."*

3.2 Estimation Results

We now document that using the LASSO to identify sparse signals in the cross-section of returns boosts out-of-sample return predictability.

OLS Model. We begin by analyzing the benchmark level of return predictability that we observe when we use a model that explicitly does not take into consideration any sparse signals. Specifically, the first column in Panel (a) of Table 1 shows the results of the predictive regressions described in Equation (4) from Subsection 2.1, which use an OLS model. We find that the average adjusted R^2 from these OLS-based return forecasts is 8.17%. That is, for a randomly selected stock, you can explain 8.17% of the variation in its minute-by-minute returns using only information about that stock's returns over the previous 3 minutes. Panel (b) of Table 1 then shows this same analysis for different subsamples of our data and indicates that the average fit is relatively stable. This predictability is not just something that exists early in our sample or during the financial crisis. If using the LASSO to account for sparse signals adds value, then it has to improve on this 8.17% benchmark.

The LASSO. Next, we examine the additional out-of-sample predictive power that comes from including the LASSO's return forecast in our model. The second column in Panel (a) of Table 1, which displays the summary statistics from the combined regressions as described in Equation (8) from Subsection 2.2, reveals that including information from the LASSO increases the out-of-sample adjusted R^2 for the typical stock by 23%, from an adjusted R^2 of 8.17% to an adjusted R^2 of 10.05%. Note that adding the LASSO's return forecast does not have to increase the adjusted R^2. First, we are studying the adjusted R^2, which takes into consideration the number of right-hand-side variables used in the model. Adding in a meaningless right-hand-side variable will actually lower the adjusted R^2 of a model. Second, we are looking at the out-of-sample fit of the LASSO.

Is this a statistically significant jump in prediction accuracy? We answer this question in two different ways. First, for each stock-month we define $\Delta_{\text{Adj. } R^2}$ as the difference between the adjusted R^2 when using both the LASSO and the OLS return forecasts and the adjusted R^2 when using only the OLS forecast. When $\Delta_{\text{Adj. } R^2} \gg 0$, then the adjusted R^2 increases a lot when we include the LASSO forecast. The rows of Table 1 labeled $\langle \Delta_{\text{Adj. } R^2} \rangle$ show that, on average, $\Delta_{\text{Adj. } R^2} = 1.88\%$ and that this difference is statistically significant at the 1% level. Second, since the regression in Equation (8) nests the regression in Equation (4), we can run an F-test to see if we should include the LASSO's return forecast when predicting a stock's return in minute $(t + 1)$. The rows of Table 1 labeled "F-test" show that we can reject the null hypothesis that the LASSO's return forecast does not add any out-of-sample predictive power at the 1% level.

3.3 Robustness Checks

We now show that these general patterns hold when we slice the data in a variety of different ways.

Subperiod Analysis. Panel (b) of Table 1 shows the predictive-regression results broken down into 3 different 3-year intervals. We now further slice our results into year-specific segments to show that the gain from including the LASSO's return forecast is consistent over time. The left panel of Figure 5 shows the average adjusted R^2

Out-of-Sample Return Predictability

a) Full Sample

$\langle \tilde{a}_n \rangle_{\times 10^4}$	0.01 (0.92)	0.01 (0.92)
$\langle \tilde{b}_n \rangle_{\times 10^4}$	4.45 (13.76)	3.87 (14.72)
$\langle \tilde{c}_n \rangle_{\times 10^4}$		1.96 (13.23)
$\langle$Adj. $R^2\rangle$	8.17%	10.05%
$\langle \Delta_{\text{Adj. } R^2} \rangle$	1.88 (8.75)	
$\langle F\text{-test} \rangle$	10.63	

b) Subperiod Analysis

	2005-07		2008-10		2011-13	
$\langle \tilde{a}_n \rangle_{\times 10^4}$	0.01 (0.73)	0.01 (0.73)	0.00 (0.22)	0.00 (0.22)	0.02 (1.29)	0.02 (1.29)
$\langle \tilde{b}_n \rangle_{\times 10^4}$	3.46 (7.14)	3.08 (7.55)	5.08 (28.02)	4.38 (23.12)	4.92 (17.34)	4.25 (19.69)
$\langle \tilde{c}_n \rangle_{\times 10^4}$		1.61 (5.87)		2.27 (37.16)		2.06 (8.80)
$\langle$Adj. $R^2\rangle$	6.41%	7.99%	9.01%	11.26%	9.31%	11.17%
$\langle \Delta_{\text{Adj. } R^2} \rangle$	1.57 (6.18)		2.25 (3.89)		1.86 (7.35)	
$\langle F\text{-test} \rangle$	9.59		11.70		10.78	

Table 1: *Average parameter estimates from the out-of-sample regressions each month using predictions from the benchmark OLS model in Equation* (4) *and the combined predictions from both the OLS model and the LASSO in Equation* (8)*. Full sample includes results from (stock,minute)-level regressions in each October from 2005 to 2013. The notation* $\langle\cdot\rangle$ *denotes the average across (permno, month) pairs. Coefficient estimates have units of percent per minute, scaled up by* 10^4 *to make the coefficients legible. Numbers in parentheses are the t-statistics, which are clustered by* (permno, month)*.* $\Delta_{\text{Adj. } R^2}$ *is the difference between the adjusted* R^2 *when using both the LASSO and OLS return forecasts and the adjusted* R^2 *when using only the OLS return forecast. "F-test" reports the results of a test that the true model for predicting out-of-sample returns does not include the LASSO's return forecast. Reads: "Including both the OLS and the LASSO forecasts increases out-of-sample return predictability by* 23%*, from an adjusted* R^2 *of* 8.17% *to an adjusted* R^2 *of* 10.05%*."*

Adjusted-R^2 Distribution, by Year

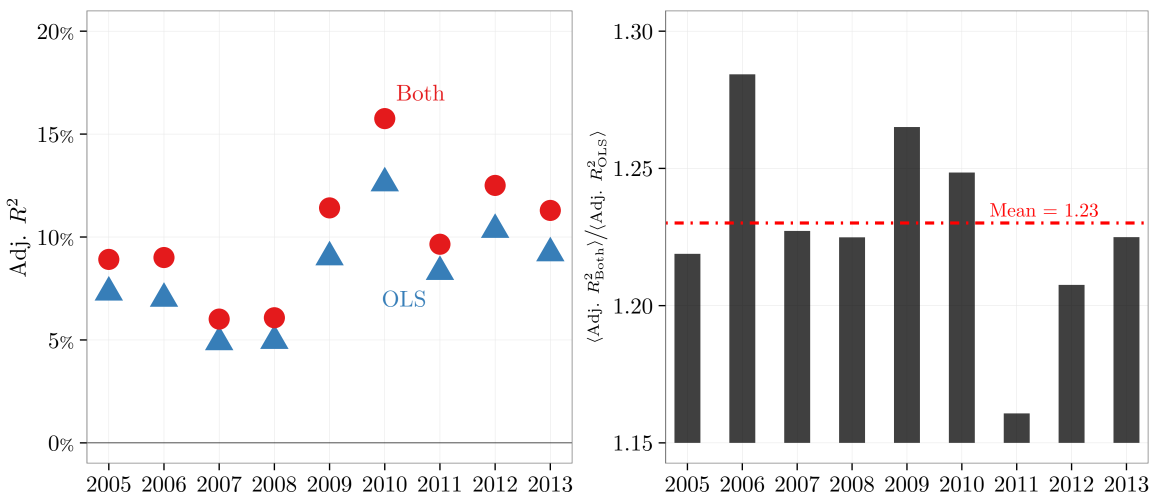

Figure 5: *Average fit from the out-of-sample regressions each month sorted by year. Left panel, blue triangles: Adjusted R^2s from predictions made using an OLS model as in Equation (4). Left panel, red circles: Adjusted R^2s from predictions made using both OLS model and the LASSO as in Equation (8). Right panel: Ratio of the average adjusted R^2 using both the OLS and the LASSO forecasts to the average adjusted R^2 using only the OLS forecast in each year. Reads: "There are some years, such as 2007, where the OLS model does a relatively poor job of forecasting returns, and there are other years, such as 2010, where the OLS model does a relatively good job of forecasting returns. But, regardless of the average fit for the OLS model in any given year, including the LASSO's return forecast always boosts out-of-sample predictive power by somewhere between 15% and 30%."*

from the predictive regressions in Equations (4) and (8) each year. The blue triangles represent the out-of-sample fit of the OLS model in each year, and the red circles represent the out-of-sample fit of the model with both the OLS and the LASSO's return forecast in each year. There are some years, such as 2007, where the OLS model does a relatively poor job of forecasting returns (blue triangle, adjusted $R^2 = 4.90\%$), and there are other years, such as 2010, where the OLS model does a relatively good job of forecasting returns (blue triangle, adjusted $R^2 = 12.62\%$).

But, notice that, regardless of the average fit for the benchmark OLS model in any given year, adding the LASSO's return forecast always boosts the out-of-sample fit of the model. The right panel of Figure 5 shows that including the LASSO's return forecast always increases out-of-sample predictive power by somewhere between 15% and 30%. What's more, the average increase in out-of-sample fit that comes with adding the LASSO's return forecast is largely unrelated to the fit of the benchmark

OLS model. In 2007, when the benchmark OLS model fit relatively poorly, adding the LASSO still bumped up the adjusted R^2 by 22%.

Industry Groupings. Next, motivated by the evidence of industry lead-lag effects documented in Hong et al. (2007), we show that the out-of-sample gain from including the LASSO's return forecast in our predictive regressions is unchanged when we slice the data by industry. We classify each stock in our sample according to its 3-digit SIC code. Figure 6 displays the increase in average adjusted R^2 values from including the LASSO's return forecast for each 3-digit industry, restricting the sample to the industries with at least 100 stocks over the course of our entire sample period. The figure shows that this 23% boost in out-of-sample return predictability from including the LASSO's return forecast is remarkably steady across all major industry groups. There are a few industries, such as metal mining, where the LASSO fits particularly well. However, there are no large industries where using the LASSO is particularly unhelpful.

Alternative Benchmarks. But, maybe the lagged returns of the same stock represents a bad choice for forecasting variables. For example, we know that at the daily and monthly horizons, the return on the market portfolio forecasts individual stock returns. We consider three alternative specifications for the OLS benchmark to show that these concerns are not driving our results. First, we include the return on the market portfolio over the previous 3 minutes as a predictor when forecasting a stock's returns using an OLS regression using Equation (6). The first column of Panel (a) in Table 2 presents the out-of-sample fit of the OLS model when using the market return. Including the market's return over the previous 3 minutes when estimating the benchmark OLS model actually lowers the out-of-sample predictive power of the benchmark model.

This lower out-of-sample predictive power suggests that the market return is a superfluous variable when trying to predict an individual stock's returns at the one-minute horizon. Adding more variables to the benchmark model will clearly raise the R^2 of this model in sample, but an additional superfluous variable will lower the model's adjusted R^2 out-of-sample due to overfitting bias. Supporting this idea, Table 7 in Appendix C shows that including the market return in the benchmark OLS specification lowers the benchmark out-of-sample predictive power by just as much as if we included a variable that is randomly drawn from standard normal distribution

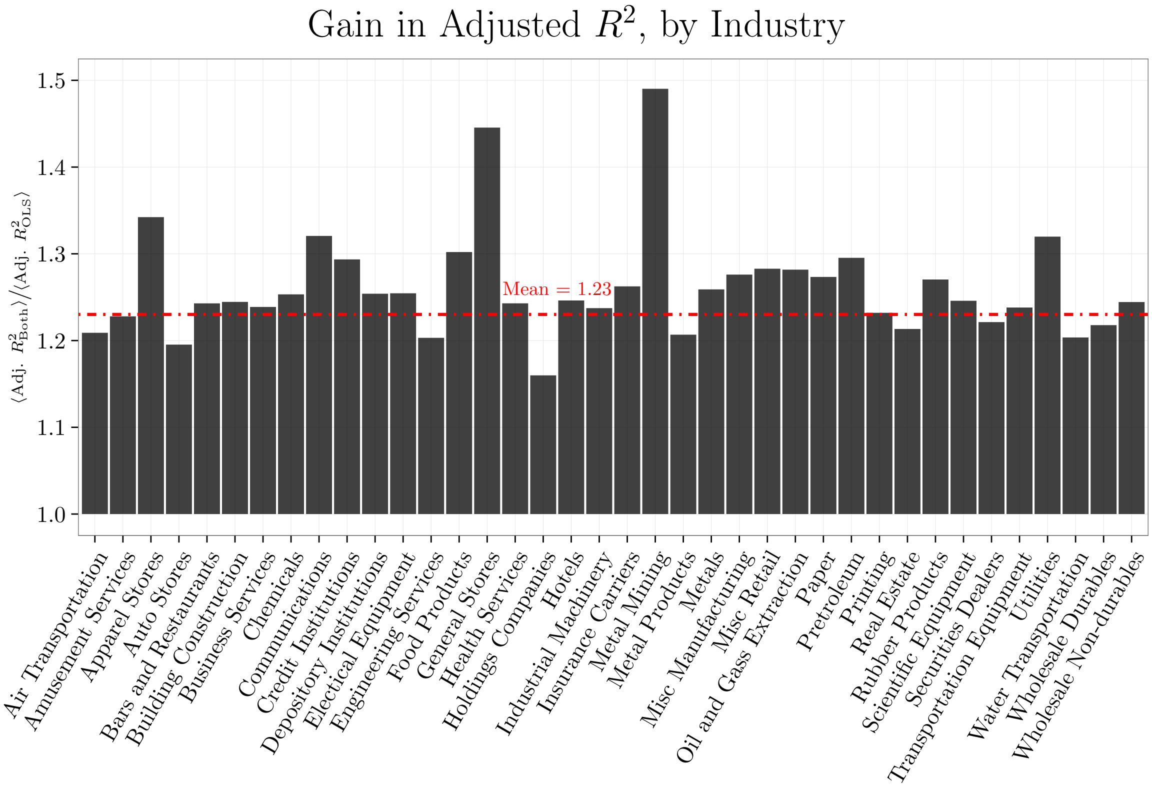

Figure 6: *Average gain from including the LASSO's return forecast when running the out-of-sample regressions each month sorted by* 3*-digit SIC-code industry. y-axis: Ratio of the average adjusted* R^2 *using both the OLS and the LASSO forecasts to the average adjusted* R^2 *using only the OLS forecast for each industry. For brevity, we only report industries with at least* 100 *observations. Reads: "The LASSO's return forecast adds significant predictive power to out-of-sample regressions in all major industries."*

in the benchmark OLS specification.

Second, when estimating the benchmark model for the nth stock, we include the return on the NYSE-listed stock with the highest pairwise correlation over the previous day. Let d_t denote the trading day on which minute t occurs. Then, for each stock minute, (n, t), we estimate $(N-1)$ regressions using the minute-by-minute data from the previous trading day,

$$\left(\tfrac{r_{n',\tau} - \hat{\mu}_{n',d-1}}{\hat{\sigma}_{n',d-1}}\right) = \tilde{\rho}_{n'} \cdot \left(\tfrac{r_{n,\tau-1} - \hat{\mu}_{n,d-1}}{\hat{\sigma}_{n,d-1}}\right) + \epsilon_{n',\tau} \quad \text{for } \tau \in (d-1), n' \neq n,$$

and choose the stock with the highest slope coefficient,

$$n_t^{\star} = \arg\max_{n' \neq n} |\tilde{\rho}_{n'}|.$$

We then include this stock's returns over the previous 3 minutes as a predictor when forecasting stock n's returns in minute t. Panel (b) of Table 2 presents the out-of-sample fit of the OLS model when including each stock's strongest pairwise predictor. Just as before, including this additional predictive variable actually lowers the out-of-sample fit of the benchmark model, from an adjusted R^2 of 8.17% to an adjusted R^2 of 4.32%. And, including this strongest pairwise predictor does not eliminate the additional out-of-sample predictive power contributed by the LASSO.

This result is worth emphasizing because it sheds light on how using the LASSO differs from using an OLS regression. Specifically, we are interested in whether or not the LASSO boosts out-of-sample forecasting power by quickly identifying sparse predictors in the cross-section of returns. Since the LASSO increases out-of-sample predictive power relative to the benchmark model with the strongest pairwise predictor over the previous day, then it must be identifying new predictors in a shorter time period. This is the role that the absolute-value penalty term in the LASSO specification plays.

Third, one might think that the benchmark OLS model would perform better if we used more lags. Again, this conjecture would certainly be true if we were only interested in the in-sample fit, but it is not necessarily true when we look at the out-of-sample fit. Using additional autoregressive lags opens up the benchmark model to overfitting bias, and this bias can be severe. In point of fact, the left panel of Figure 7 reveals that the corrected Akaike information criterion (AICc) chooses 3 or fewer lags in more than 82% of the 30-minute time windows that we study. In keeping with the idea that using a small number of lags is optimal, Panel (c) of Table 2 shows the out-of-sample fit of the benchmark OLS model when including the optimal number of lags according to the AICc. There is little difference between the performance of this optimizes autoregressive model and the autoregressive model with 3 lags. If anything, using an autoregressive model with an AICc-optimal number of lags performs slightly worse than using a simple AR(3) model. Table 6 in Appendix C shows the performance of the LASSO relative to benchmark OLS models with 1, 2, 4, and 5 lags. In all of these specifications the LASSO significantly boosts out-of-sample predictive power relative to the benchmark model. In summary, our results are not driven by the particular details of how we define our benchmark model.

Penalty Parameters. Finally, we find that our results are robust to selecting the penalty parameter, λ, in different ways. We select the λ for each stock by choosing the penalty parameter with the highest out-of-sample R^2 during the first 45 minutes

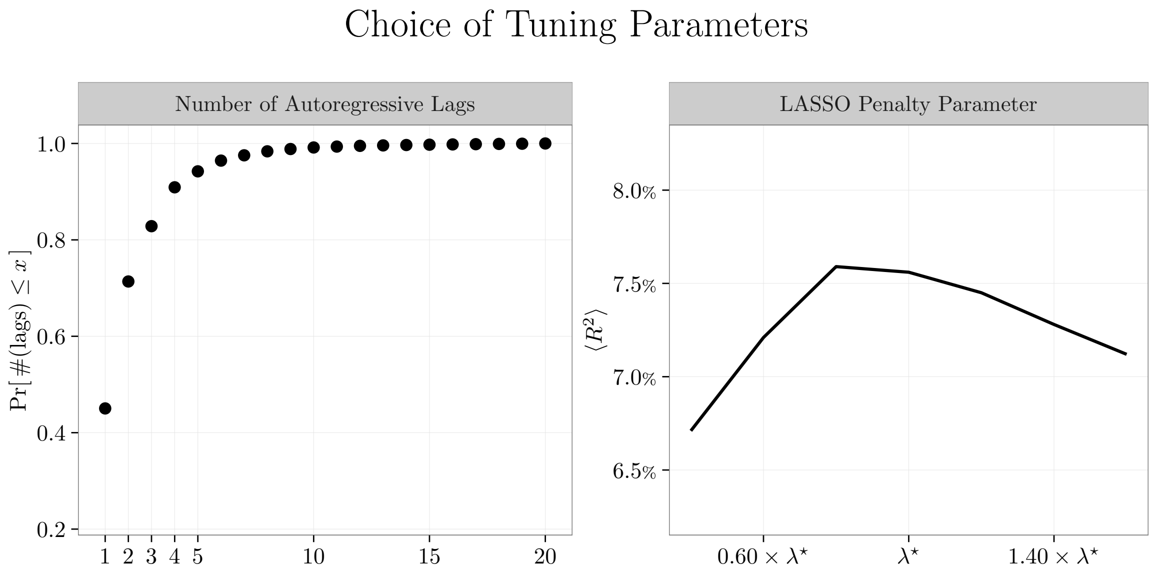

Figure 7: *Left panel: Fraction of the different* 30*-minute time windows where the OLS model in Equation* (3) *chooses a given number of lags according to the Akaike information criterion (AIC). Reads: "There are more than* 3 *informative lags in less than* 18% *of the* 30*-minute time windows." Right panel: Average out-of-sample adjusted* R^2 *for the LASSO prediction when using non-optimized penalty parameters for a random sample of* 20 *stocks. Reads: "When the LASSO is estimated each day with a penalty parameter that is* 60% *of the optimal choice, the resulting average adjusted* R^2 *for the prediction is* 7.24%*."*

of each trading day. This parameter then remains constant throughout the rest of the trading day. Choosing the penalty parameter using the first 45 minutes of the trading day is not necessarily optimal, but we simply want to show that accounting for sparse signals can significantly boost out-of-sample predictive power. This procedure is the method of choice in Friedman et al. (2010). The right panel of Figure 7 shows that the LASSO's predictions do not depend on the gritty details of how λ is chosen. Our results persist if we use a LASSO penalty parameter that is within $\pm 40\%$ of the optimal λ as estimated during the first 45 minutes of the trading day.

3.4 Evidence of Sparsity

In the previous subsection, we saw that the LASSO must be identifying new predictors in a shorter time period than standard OLS models because the LASSO increases out-of-sample predictive power relative to benchmark models that include the strongest pairwise predictors over the previous day. We now dig into this result even deeper and verify that LASSO's out-of-sample predictive power comes from quickly identifying

Out-of-Sample Return Predictability, Alternative Benchmarks

	Market Returns		Pairwise Predictor		Optimal # Lags	
$\langle \tilde{a}_n \rangle_{\times 10^{-4}}$	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)
$\langle \tilde{b}_n \rangle_{\times 10^{-4}}$	3.98 (13.49)	3.36 (14.51)	3.20 (13.03)	2.60 (14.40)	3.57 (13.05)	3.00 (14.22)
$\langle \tilde{c}_n \rangle_{\times 10^{-4}}$		2.14 (12.64)		2.53 (12.56)		2.40 (12.81)
$\langle$Adj. $R^2\rangle$	6.65%	8.80%	4.32%	7.21%	5.43%	8.08%
$\langle \Delta_{\text{Adj. } R^2} \rangle$	2.16 (7.76)		2.89 (7.72)		2.65 (8.05)	
$\langle F\text{-test} \rangle$	11.35		13.17		12.59	

Table 2: *Average of the parameter estimates from the out-of-sample regressions each month described by Equations (4) and (8) with alternative specifications for the benchmark OLS model. Market Returns: benchmark model for stock n includes 3 lagged returns for stock n and 3 lags of the market return. Pairwise Predictor: benchmark model for stock n includes 3 lagged returns for stock n and 3 lagged returns of the stock with the highest pairwise correlation with stock n in the previous trading day. Optimal # Lags: benchmark model for stock n includes the optimal number of lagged returns for stock n according to the corrected Akaike information criterion. Full sample includes results from (stock,minute)-level regressions in each October from 2005 to 2013. Coefficient estimates have units of percent per minute. Numbers in parentheses are the t-statistics, which are clustered by (permno, month). $\Delta_{Adj.\ R^2}$ is the difference between the adjusted R^2 when using both the LASSO and OLS return forecasts and the adjusted R^2 when using only the OLS return forecast. "F-test" reports the results of a test that the true model for predicting out-of-sample returns does not include the LASSO's return forecast.*

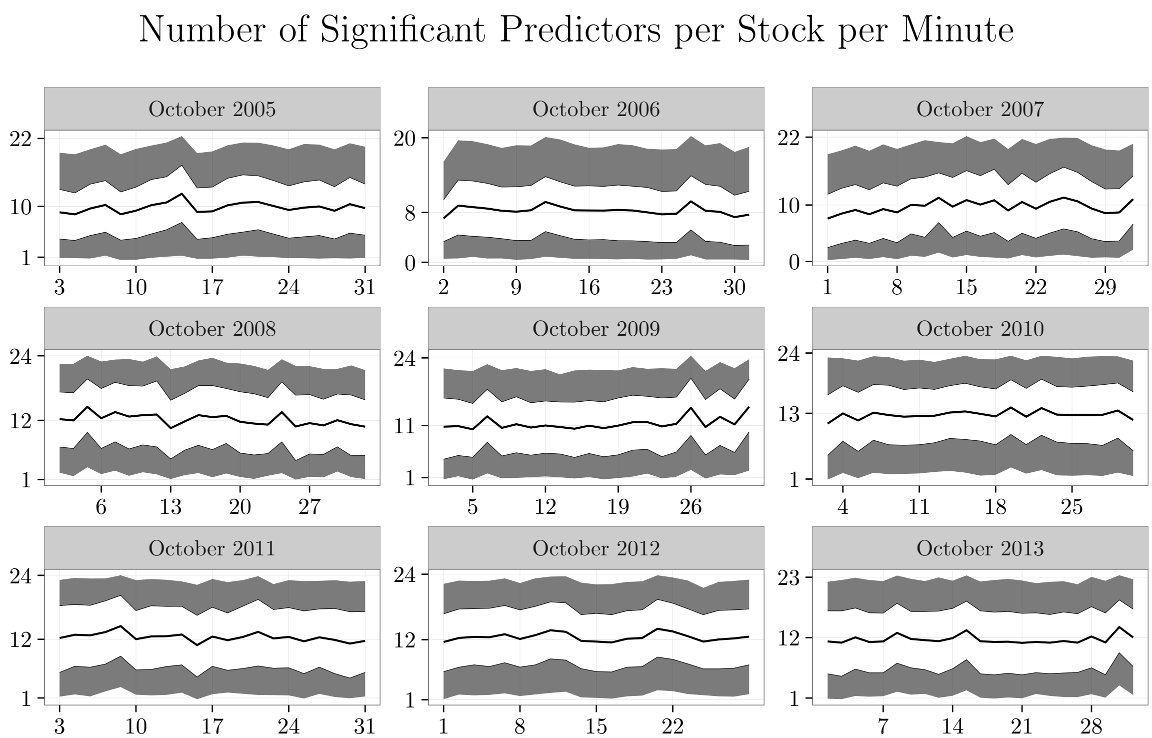

Figure 8: *Number of predictors used by the LASSO to makes its* 1*-minute-ahead return forecasts. Solid black line: Daily average number of significant predictors selected by the LASSO each minute. Grey ribbons:* $(5\%, 25\%]$ *and* $[75\%, 95\%)$ *ranges for the number of significant predictors selected by the LASSO each minute. Days labeled on the* x*-axis are Mondays. The values labeled on the* y*-axis correspond to the minimum, mean, and maximum of the number of predictors used by the LASSO each the month. Data: NYSE-listed stocks traded in each October from 2005 to 2013. Reads: "If you select a stock at random and then look at the LASSO's return forecast for that stock in a randomly selected minute during October 2010, then you should expect the LASSO to use roughly* 13 *predictors when making this return forecast."*

the right variables at the right time, not from better estimating the effect of some persistent factor.

Number of Predictors. To start with, the LASSO uses an extremely small number of predictors to make its return forecast every minute. Figure 8 characterizes the number of significant predictors the LASSO uses to make its return forecast for each NYSE-listed stock in each minute. On average, the LASSO uses only 11 predictors to make its return forecast. To put this number in perspective, note that this is roughly

$$0.55\% = {}^{11}\!/_{2{,}000}$$

of the roughly 2,000 possible stocks that the LASSO could choose from each month. Moreover, this is a stable feature of all the stocks we look at. The LASSO uses the smallest number of predictors on average in October 2006—only 8—and the largest number of predictors on average in October 2010—over 13; but, this range is still exceptionally small relative to the total number of stocks—roughly 2,000.

The LASSO's tendency to use only a handful of predictors is also extremely stable over time. The thick black line gives the average number of significant predictors selected by the LASSO in each minute; whereas, the grey shaded regions give the $(5\%, 25\%]$ and $[75\%, 95\%)$ ranges. While the LASSO does tend to use slightly more predictors later in the sample period, the basic pattern is constant across our sample. There are no weekly patterns in the number of predictors used. The cross-sectional distribution of the number of predictors used is pretty stable over time.

Predictor Turnover. Even though the LASSO uses a small number of predictors when making its return forecasts, this is not necessarily evidence that it is identifying a sparse signal. It could be the case that the LASSO always chooses the same 11 predictors—in other words, that the signal is just a persistent factor. However, the LASSO's chosen predictors do not remain significant for long. Figure 9 shows that the median predictor emerges into significance for a single minute, sees its shadow, and then disappears. Moreover, less than 10% of all LASSO predictors remain significant for more than 4 minutes. If we estimate a simple hazard-rate model, we find that each significant predictor has a 60.4% chance of becoming insignificant in the following minute. This means that, if the LASSO uses 11 different predictors to make its return forecast in the current minute, then on average the LASSO will not be using any of these predictors 8 minutes later because the expected time until 11 failures is given by

$$7.33_{\,\mathrm{min}} = \left(\tfrac{1-0.60}{0.60}\right) \times 11$$

according to a simple negative-binomial model.

4 Economic Origins

This section examines the economic origins of the LASSO's out-of-sample predictive power. We find that the LASSO identifies information about real-world events that would be difficult to uncover using intuition alone, like in the Family Dollar example.

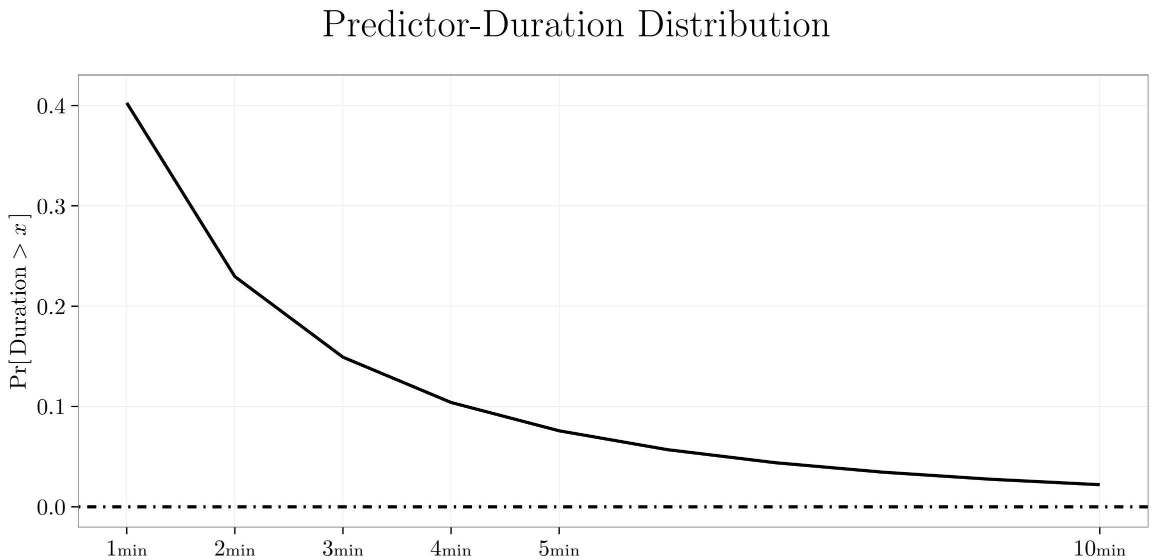

Figure 9: *Probability that the LASSO uses a predictor for more than x consecutive minutes when making the return forecast for a particular stock. Reads: "Less than* 10% *of all LASSO predictors remain significant for more than* 4 *consecutive minutes."*

First, we document that, even though we fit the LASSO separately when making return forecasts for different stocks, these separate forecasts still tend to identify the same predictors at the same time. Then, we show that this predictor clustering is related to subsequent news announcements. The lagged returns of stocks with announcements are 18.3% more likely to be used by the LASSO as a predictor during the previous 30 minutes.

4.1 Predictor Clustering

We begin the analysis in the this section by showing that, even though we fit the LASSO separately when making return forecasts for different stocks, these separate forecasts still tend to identify the same predictors at the same time far more often than would be expected by pure chance.

Random Overlap. How often should we expect the LASSO to choose the same predictors when forecasting the returns of different stocks by pure chance? We construct this benchmark via numerical simulations. In each minute during our sample, we compute the average number of predictors selected by the LASSO when forecasting stock returns, q_t. To provide some context, recall that, over the course of the

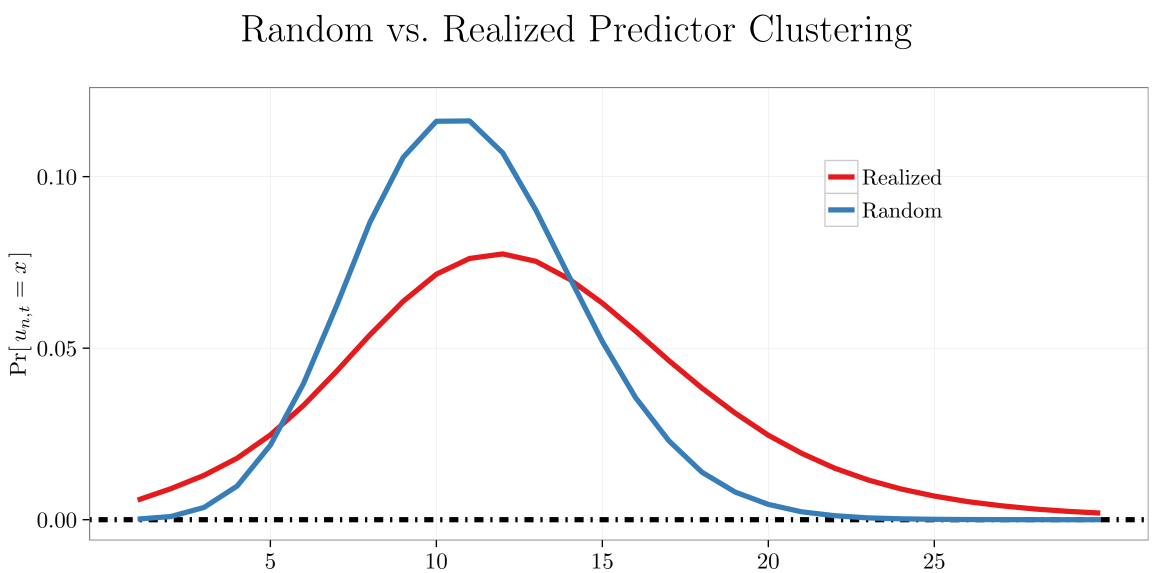

Figure 10: *Blue line: Probability that a particular predictor would be used to make return forecasts for x different stocks if each stock's predictors were chosen at random as defined in Equation (9). Red line: Probability that the LASSO uses a particular predictor when making its return forecast for x different stocks as defined in Equation (10). Reads: "The LASSO is* 17.11*-times more likely to use a predictor in more than* 20 *of its return forecasts than would be expected by pure chance."*

entire sample period, the LASSO uses the lagged returns of $\langle q_t \rangle = 11$ other stocks when making its return forecasts. To create our benchmark level of predictor overlap, we then randomly select q_t predictors for each stock, $n = 1, 2, \ldots, N$. Let $\mathcal{Q}_{n,t}^{\text{Random}}$ denote the set of q_t predictors randomly chosen for the nth stock in minute t.

Let $u_{n,t}^{\text{Random}}$ denote the number of times that stock n is randomly chosen as a predictor for some stock $n' \in \{1, 2, \ldots, N\}$ in minute t,

$$u_{n,t}^{\text{Random}} = \textstyle\sum_{n'=1}^{N} 1\{n \in \mathcal{Q}_{n',t}^{\text{Random}}\}. \tag{9}$$

That is, to create our benchmark measure of predictor usage, $u_{n,t}^{\text{Random}}$, for the nth stock in minute t we look at the group of $\mathcal{Q}_{n',t}^{\text{Random}}$ stocks that we randomly selected for each stock $n' \in N$ for minute t in a given simulation and count up the number of times that stock n happens to be in this set. We estimate the distribution of $u_{n,t}^{\text{Random}}$ when the choice of predictors is randomly selected using 10^5 iterations in each minute and plot the results using the blue line in Figure 10. If predictors were just randomly selected each minute, then there would be very few stocks used in a large number of LASSO predictions in any given minute.

Realized Overlap. Compared to this random benchmark, the LASSO's actual choice of predictors each minute is far more likely to cluster on a few stocks. Let $\mathcal{Q}_{n,t}$ denote the set of predictors that the LASSO actually chooses to use when forecasting the returns of the nth stock in minute t. We can then define a measure of the nth stock's actual usage in minute t as

$$u_{n,t} = \sum_{n'=1}^{N} 1\{n \in \mathcal{Q}_{n',t}\}. \tag{10}$$

That is, to measure the LASSO's actual usage of the nth stock in minute t, $u_{n,t}$, we look at the group of stocks that the LASSO chose to use, $\mathcal{Q}_{n',t}$, when making its forecast for each NYSE-listed stock $n' \in N$ in minute t and count up the number of times that stock n happens to be in this set. The red line in Figure 10 shows the probability that, in a given minute, some stock was chosen as a predictor by the LASSO x times. It is clear from Figure 10 that separate LASSO forecasts select the same predictors far more often than would be expected by pure chance.

Predictor Clustering. We estimate that, compared to the random-selection benchmark, the LASSO is 17.11-times more likely to use a predictor in more than 20 of its return forecasts. In other words, the LASSO's choice of predictors is much more likely to cluster on a handful of selections than would be expected if the LASSO's choices of predictors for different stocks were unrelated. We now show that is predictor clustering this associated with subsequent news announcements.

4.2 RavenPack Data

We obtain business-press data from RavenPack to investigate how the LASSO's choice of predictors is related to news announcements.

Data Source. RavenPack has a partnership with Dow Jones, giving it access to the full Dow Jones news archives. These data consist of all Dow Jones Newswire and Wall Street Journal articles. The Dow Jones news archives have been used in many prior studies (Kolasinski et al., 2013; Shroff et al., 2013; Dai et al., 2014). The RavenPack data is particularly useful for our analysis because it occurs at a frequency faster than the one-minute level that our return data occur at. So, we can use it to examine the relative timing of news announcements and predictor usage.

Variable Definitions. We compute a "has news" indicator variable for each stock in each minute of our sample,

$$hasNews_{n,t} = \begin{cases} 1 & \text{if there is a new release about stock } n \text{ in minute } t, \\ 0 & \text{else.} \end{cases} \tag{11}$$

We then create a 60-minute event-time window around each (stock, minute) with a news announcement where time is indexed by $h \in \{-30, \ldots, -1, 0, 1, \ldots, 29\}$. If the LASSO tends to select stocks as predictors in the minutes before a news announcement, then $u_{n,h}$ should be higher for minutes $h \in \{-30, \ldots, -1\}$. So, we create an indicator variable, $before_{n,h}$, which equals one during the 30 minutes before a news announcement and zero otherwise.

One might be concerned that news announcements cluster in time. So, it might be difficult to tell whether the LASSO is selecting a stock as a predictor in the 30 minutes leading up to one news announcement or in the 30 minutes following an earlier announcement. To address this concern, we create an indicator variable, $earlierEvent_{n,h}$, that equals one if there was another news announcement about stock n during the previous 30 minutes and zero otherwise. By interacting the "earlier-event" indicator variable with the "before-a-news-announcement" indicator variable, we can tease out whether the observed effects are due to future news announcements

Excess Predictor Clustering

	$\Pr(x > 15)$		$\Pr(x > 20)$		$\Pr(x > 25)$	
	$u_{n,t}^{\text{Random}}$	$u_{n,t}$	$u_{n,t}^{\text{Random}}$	$u_{n,t}$	$u_{n,t}^{\text{Random}}$	$u_{n,t}$
Mean	0.0920	0.2742	0.0046	0.0787	0.0001	0.0169
Ratio	2.98		17.11		199.00	

Table 3: $u_{n,t}^{Random}$*: number of times a stock is selected at random as a predictor as defined in Equation (9).* $u_{n,t}$*: number of times a stock is selected by the LASSO as a predictor as defined in Equation (10).* $\Pr(x > 15)$*: columns contain the probabilities that a stock is used as a predictor at least* 15 *times in a single minute.* $\Pr(x > 20)$*: columns contain the probabilities that a stock is used as a predictor at least* 20 *times in a single minute.* $\Pr(x > 25)$*: columns contain the probabilities that a stock is used as a predictor at least* 25 *times in a single minute. Rows labeled "Mean" report the mean probabilities for a given threshold. Rows labeled "Ratio" report the ratio of these means for a given threshold. Reads: "When predictors are chosen randomly each minute, there is only a* 0.46% *chance that a single stock is selected more than* 20 *times in a given minute. Compared to this random-selection benchmark, the LASSO is* $^{0.0787}/_{0.0046} = 17.11$*-times more likely to use a predictor in more than* 20 *of its return forecasts in a given minute."*

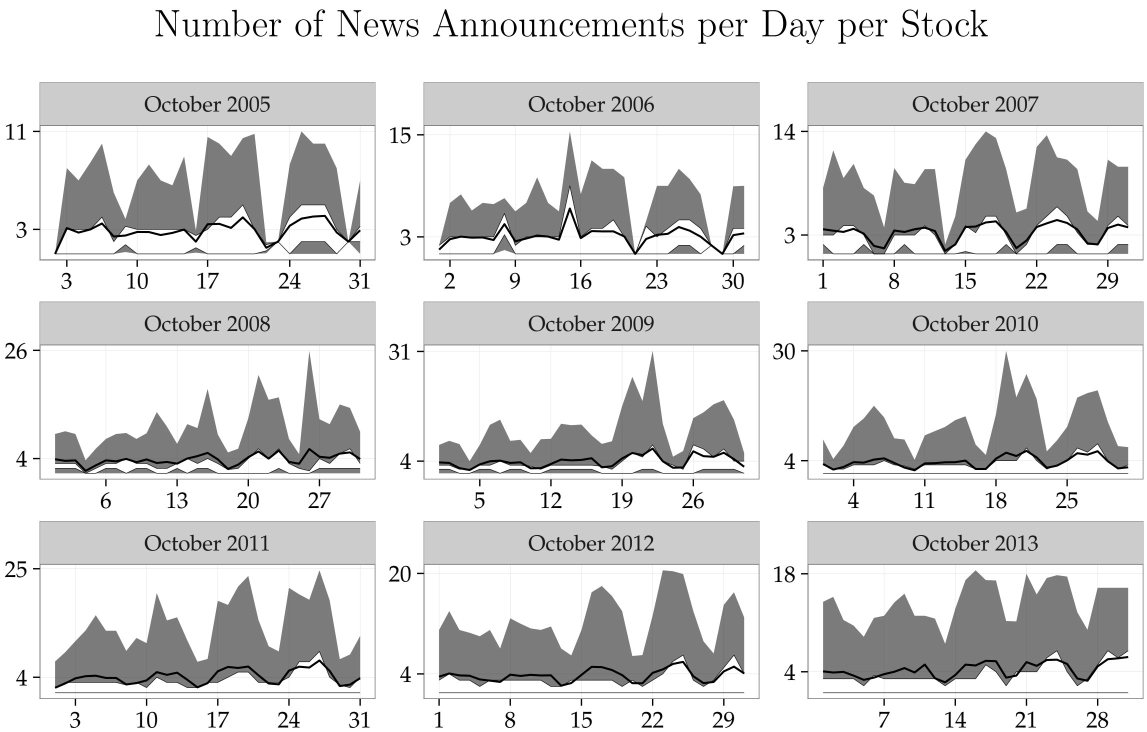

Figure 11: *Distribution of the number of news releases for each stock each day in the RavenPack data. Thick black line: Daily average number of news releases across all NYSE-listed stocks. Grey ribbon:* $[75\%, 95\%)$ *range for the number of news releases per stock. Thin black line: Number of news releases for the NYSE-listed stock with the most news. Days labeled on the x-axis are Mondays. The values labeled on the y-axis correspond to the mean, and maximum of the number of news releases per day for each stock during the month. Data: RavenPack news-release data for NYSE-listed stocks traded in each October from 2005 to 2013. Reads: "During October 2009, the typical NYSE-listed stock had* 4 *news releases per day, while the stock with the most news had* 159*."*

or to effects of news-announcement clustering.

After collecting each news release, RavenPack assigns it both a relevance score ranging from 0.0 (not relevant) to 1.0 (most relevant) and a sentiment score ranging from 0.0 (pessimistic) to 1.0 (optimistic). We normalize these variables to have a mean of 0 and a standard deviation of 1. In our regression analysis, for example, we can then interact the relevance variable with our "before-a-news-announcement" indicator variable to investigate whether the LASSO is more likely to select a stock as a predictor in the minutes before a major negative news announcement.

Summary Statistics. Panel (a) of Table 4 shows the summary statistics for the RavenPack data. The first row reports that, in the 60-minute windows around

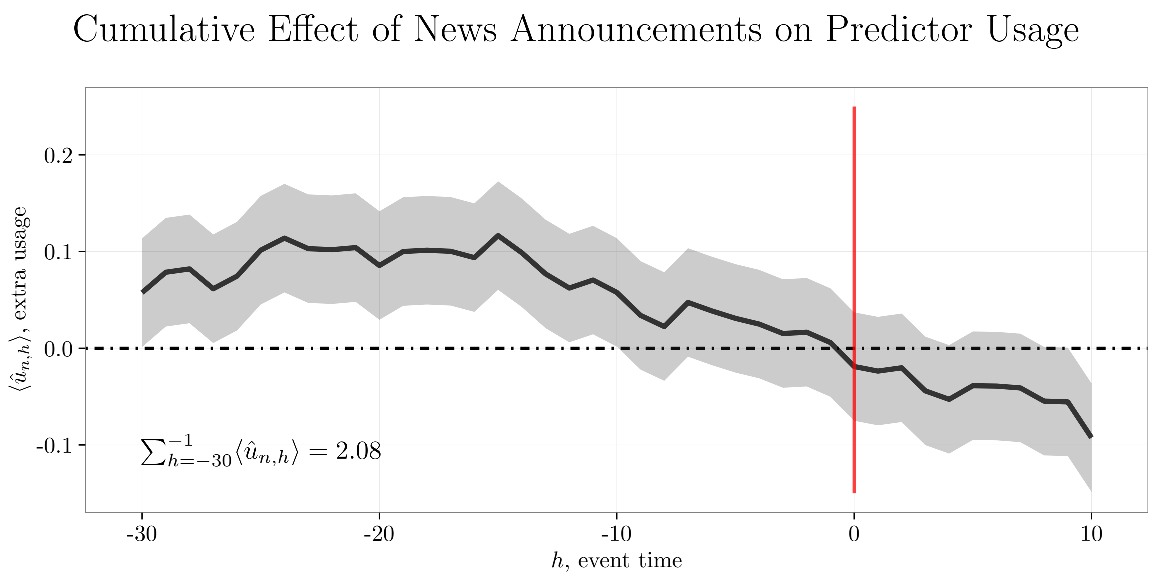

Figure 12: *Black line: additional number of times stock n is used as a predictor by the LASSO in event-time minute h, $\langle \hat{u}_{n,h} \rangle$, as defined in Equation (13). Grey band:* 95% *confidence interval for this point estimate in each event-time minute. Data: RavenPack news-release data for NYSE-listed stocks traded in each October from 2005 to 2013. Reads: "Over the course of the* 30 *minutes leading up to a news announcement, the LASSO chooses a stock as a predictor an additional* $\sum_{h=-30}^{-1} \langle \hat{u}_{n,h} \rangle = 2.08$ *times."*

news announcements, stocks are used by the LASSO as a predictor 11.42 times per minute—that is, roughly the same number of times as in the sample average. The second row reveals that news announcements exhibit significant clustering. For roughly 53% of all news announcements, there is another announcement in the previous 30 minutes. Finally, the third and fourth rows document the summary statistics for the relevance and sentiment variables.

4.3 Event-Study Analysis

We use an event-study methodology to show that the LASSO is more 18.3% more likely to use a stock as a predictor in the minutes leading up to a news announcement.

Baseline Specification. In our baseline specification, we look at the 60-minute window around the (stock, minute) where a news announcement takes place, $h \in \{-30, \ldots, -1, 0, 1, \ldots, 29\}$, and ask whether or not a stock is used more often by the LASSO as a predictor in the 30 minutes prior to a news announcement,

$$u_{n,h} = a + b \cdot \mathit{before}_{n,h} + e_{n,h}. \tag{12}$$

News-Announcement Event Study

a) Summary Statistics

	μ	σ	min	q_{25}	q_{50}	q_{75}	max
$u_{n,h}$	11.42	7.96	0.00	6.00	11.00	15.00	289.00
$earlierEvent_{n,h}$	0.53	0.50					
$relevance_{n,h}$	0.00	1.00	−5.50	0.14	0.27	0.58	0.66
$sentiment_{n,h}$	0.00	1.00	−9.54	0.00	0.00	0.36	10.08

b) Regression Analysis

	Dependent Variable = $u_{n,h}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
const.	11.39 (72.27)						
$before_{n,h}$	0.11 (2.63)	0.11 (2.63)	0.11 (2.66)	0.11 (2.65)	0.09 (2.41)	0.11 (2.69)	0.11 (2.63)
$earlierEvent_{n,h}$					0.21 (3.45)		
$relevance_{n,h}$						−0.09 (4.42)	
$sentiment_{n,h}$							−0.02 (0.84)
$before_{n,h} \times earlierEvent_{n,h}$					0.05 (0.65)		
$before_{n,h} \times relevance_{n,h}$						0.09 (2.64)	
$before_{n,h} \times sentiment_{n,h}$							−0.02 (0.63)
Calendar Time FE		✓		✓	✓	✓	✓
Stock FE			✓	✓	✓	✓	✓

Table 4: *Panel a) Summary statistics for variables used in Equations* (12), (14), (15), *and* (16). *Panel b) Coefficient estimates from the regressions in Equations* (12), (14), (15), *and* (16). *The first* 3 *columns use stock-by-minute-level data during trading hours during each October from 2005 to 2013. The fourth column uses ordered-stock-pair-by-minute-level data over the same sample period. All* 4 *regressions have time and group fixed effects. Coefficient estimates have units of predictors. Numbers in parentheses are the t-statistics. Reads: "The LASSO typically uses a stock with a news announcement in* 0.11 *additional predictions."*

If the estimated coefficient b is greater than 0, then the LASSO is more likely to use a stock as a predictor in the minutes before a news announcement. This is exactly what we find in the first column of Panel (b) in Table 4. A stock is used an additional 0.11 times per minute in the minutes leading up to a news announcement.

One might be concerned that the effect is driven by stock- or minute-specific effects. For example, it could be the case that the stocks that realize the most news announcements are also used by the LASSO as predictors the most often. Or, it could be the case that the LASSO uses the largest number of predictors to make its return forecasts in the minutes (calendar time) that have the most news announcements. To address these concerns, we include stock and calendar-time fixed effects in our regressions. Let $\hat{u}_{n,h}$ denote the predictor usage of stock n in event-time minute h after accounting for these fixed effects,

$$\hat{u}_{n,h} = u_{n,h} - \mathrm{E}(u_{n,h}|n, t). \tag{13}$$

Columns 2, 3, and 4 in Panel (b) of Table 4 show that including these fixed effects does not alter the point estimate from the baseline regression. The LASSO tends to use a stock as a predictor right before it has a news announcement, regardless of which stock you look at or which subperiod you look in.

This point estimate of $b = 0.11$ is also large. Figure 12 plots the average value of $\hat{u}_{n,h}$ in each event-time minute. After controlling for stock and calendar-time fixed effects, the LASSO is much more likely to select a stock as a predictor in the minutes before a news announcement. When we sum up these average values over the 30 minutes prior to an announcement, we get a value of $\sum_{h=-30}^{-1} \langle \hat{u}_{n,h} \rangle = 2.08$—that is, over the course of the 30 minutes leading up to a news announcement, the LASSO chooses a stock as a predictor an additional 2.08 times.

Earlier Announcements. The summary statistics in Panel (a) of Table 4 suggest that news announcements cluster in time. So, the $b = 0.11$ point estimate from the baseline specification could be indicating that the LASSO selects stocks as predictors in the 30 minutes leading up to news announcements or in the 30 minutes following earlier announcements. To tease apart these hypotheses, we interact the "before-a-news-announcement" indicator variable with a "had-an-earlier-announcement" indicator variable that equals one whenever there was a news announcement in the previous

30 minutes,

$$\hat{u}_{n,h} = a + b \cdot \mathit{before}_{n,h} + c \cdot \mathit{earlierEvent}_{n,h} + d \cdot \{\mathit{before}_{n,h} \times \mathit{earlierEvent}_{n,h}\} + e_{n,h}. \quad (14)$$

The fifth column of Panel (b) in Table 4 shows the results for this regression and reveals that this $b = 0.11$ point estimate is not being driven by announcement clustering. We estimate that $c = 0.21$, which implies that the LASSO is more likely to use a stock as a predictor when it has several news announcements in quick succession. But, after accounting for earlier announcements, the coefficient on the "before-an-announcement" indicator variable is still positive and significant while the coefficient on the interaction term is statistically insignificant. This combination of coefficients suggests that the timing of the LASSO's choice of predictors is not being driven by news-announcement clustering. The LASSO is consistently picking as predictors stocks with *subsequent* news announcements.

Relevance of Information. We argue that the LASSO is using the cross-section of stock returns to identify sparse signals with real-world information content. If this is true, then the LASSO should be more likely to use a stock as a predictor when the subsequent news announcement is more relevant. To test this hypothesis, we estimate the regression below,

$$\hat{u}_{n,h} = a + b \cdot \mathit{before}_{n,h} + c \cdot \mathit{relevance}_{n,h} + d \cdot \{\mathit{before}_{n,h} \times \mathit{relevance}_{n,h}\} + e_{n,h}, \quad (15)$$

where we interact our "before-a-news-announcement" indicator variable with a variable that reflects the relevance of the news announcement at time $h = 0$ to stock n. More relevant stories have higher values, and we normalize this relevance variable to have a mean of 0 and a standard deviation of 1. The sixth column of Panel (b) in Table 4 reveals that the LASSO is much more likely to use a stock as a predictor in the minutes leading up to an extremely relevant news announcement. If an announcement is only tangentially related to stock n, then the LASSO is no more likely than chance to use stock n as a predictor.

Positive vs. Negative Announcements. Finally, we investigate whether the LASSO's predictor choice depends on the nature of the news announcement. Specifically, we look at whether positive and negative news stories have the same effect on the LASSO's predictor choice. We do this by using the same interaction-effect methodology as before,

$$\hat{u}_{n,h} = a + b \cdot \mathit{before}_{n,h} + c \cdot \mathit{sentiment}_{n,h} + d \cdot \{\mathit{before}_{n,h} \times \mathit{sentiment}_{n,h}\} + e_{n,h} \quad (16)$$

where we interact the "before-a-news-announcement" indicator variable with a variable that reflects the sentiment of the news announcement at time $h = 0$ for stock n. More positive stories have higher values, and we normalize this sentiment variable to have a mean of 0 and a standard deviation of 1. The seventh column of Panel (b) in Table 4 shows that the LASSO does not care whether the story is positive or negative. The coefficient on the sentiment variable and its interaction term are statistically indistinguishable from zero.

5 Conclusion

This paper uses the LASSO to identify rare, short-lived, "sparse" predictors in the cross-section of returns. We show that using the LASSO to identify these sparse signals boosts out-of-sample predictability in one-minute returns by 23% relative to standard OLS regressions. And, we verify that this out-of-sample predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some new, persistent factor. When we examine the economic origins of the LASSO's success, we find that it comes from identifying predictors that realize subsequent news announcements. A stock with a news announcement in minute t is 18.3% more likely to be used by the LASSO as a predictor in minutes $\{(t-30), \ldots, (t-1)\}$. Thus, we show that it is possible to use penalized regressions like the LASSO to identify information about real-world events that would be difficult to uncover using standard regression techniques. This link between LASSO predictors and subsequent news announcements provides an economic motivation for the use of machine-learning techniques, like the LASSO, to forecast returns.

References

Ahn, D., J. Conrad, and R. Dittmar (2003). Risk adjustment and trading strategies. *Review of Financial Studies 16*(2), 459–485.

Asness, C. (1994). Variables that explain stock returns. Ph.D. Dissertation, University of Chicago.

Banz, R. (1981). The relationship between the returns and the market values of common stocks. *Journal of Financial Economics 9*(1), 3–18.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*(3), 307–343.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Berk, J., R. Green, and V. Naik (1999). Optimal investment, growth options, and security returns. *Journal of Finance 54*(5), 1553–1607.

Boudoukh, J., M. P. Richardson, and R. Whitelaw (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Review of Financial Studies 7*(3), 539–573.

Budish, E., P. Crampton, and J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics 130*(4), 1547–1621.

Campbell, J. and S. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Candes, E. and T. Tao (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics 35*(6), 2313–2351.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Chan, L., N. Jegadeesh, and J. Lakonishok (1996). Momentum strategies. *Journal of Finance 51*(5), 1681–1713.

Chernozhukov, V., C. Hansen, and M. Spindler (2015). Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review 105*(5), 486–490.

Chinco, A. (2015). Feature-selection risk. Working Paper.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *Journal of Finance 63*(4), 1977–2011.

Conrad, J., G. Kaul, and M. Nimalendran (1991). Components of short-horizon individual security returns. *Journal of Financial Economics 29*(2), 365–384.

Dai, L., J. Parwada, and B. Zhang (2014). The governance role of the media through news dissemination: evidence from insider trading. Working Paper.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *Journal of Finance 53*(6), 1839–1885.

De Bondt, W. and R. Thaler (1985). Does the stock market over-react? *Journal of Finance 40*(3), 793–805.

DeMiguel, V., L. Garlappi, F. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science 55*(5), 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

DeMiguel, V., F. Nogales, and R. Uppal (2014). Stock return serial dependence and out-of-sample portfolio performance. *Review of Financial Studies 27*(4), 1031–1073.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics 32*(2), 407–499.

Fama, E. (1965). The behavior of stock-market prices. *Journal of Business 38*(1), 34–105.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software 33*(1), 1–22.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics 129*(4), 1661–1710.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (1 ed.). Springer Science & Business Media.

Hong, H. and J. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance 54*(6), 2143–2184.

Hong, H., W. Torous, and R. Valkanov (2007). Do industries lead stock markets? *Journal of Financial Economics 83*(2), 367–396.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Johnson, T. (2002). Rational momentum effects. *Journal of Finance 57*(2), 585–608.

Kolasinski, A., A. Reed, and M. Riggenberg (2013). A multiple lender approach to understanding supply and search in the equity lending market. *Journal of Finance 68*(2), 559–595.

Liu, X. and L. Zhang (2008). Momentum profits, factor pricing, and macroeconomic risk. *Review of Financial Studies 21*(6), 2417–2448.

Lo, A. and C. MacKinlay (1990). When are contrarian profits due to stock-market overreaction? *Review of Financial Studies 3*(2), 175–205.

Manela, A. and A. Moreira (2015). News-implied volatility and disaster concerns. *Journal of Financial Economics Forthcoming*.

Meinshausen, N. and B. Yu (2009). LASSO-type recovery of sparse representations for high-dimensional data. *Annals of Statistics 37*(1), 246–270.

Moskowitz, T., Y. Ooi, and L. Pedersen (2012). Time-series momentum. *Journal of Financial Economics 104*(2), 228–250.

Shroff, N., R. Verdi, and G. Yu (2013). Information environment and the investment decisions of multinational corporations. *Accounting Review 89*(2), 759–790.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B 58*(1), 267–288.

Zhu, Y. and G. Zhou (2009). Technical analysis: An asset allocation perspective on the use of moving averages. *Journal of Financial Economics 92*(3), 519–544.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B 67*(2), 301–320.

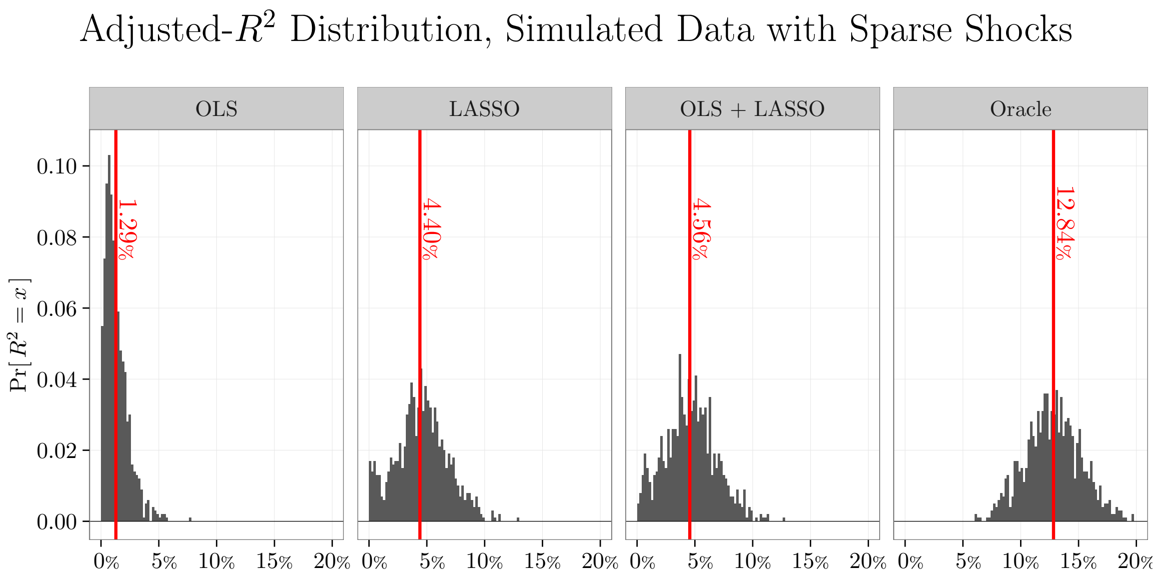

Figure 13: *Distribution of adjusted R^2s from the forecasting regressions using simulated data generated from Equation (17). Black bars: Probability that the adjusted R^2 from a single out-of-sample forecasting regression falls within a 0.1%-point interval. Red vertical line: Average adjusted R^2 from these regressions. Far-left panel: Out-of-sample prediction made using OLS. Left-center panel: Out-of-sample prediction made using the LASSO. Right-center panel: Out-of-sample predictions made using both OLS and the LASSO. Far-right panel: Out-of-sample prediction made using OLS but including the true set of $K = 5$ predictors as right-hand-side variables. Reads: "Including the LASSO's return forecast boosts the out-of-sample adjusted R^2 from 1.29% to 4.56% in simulated data."*

A Simulation Analysis

We apply the LASSO to simulated returns in order to verify that it really is identifying sparse signals in the cross-section of returns. All of the relevant code is available in an online appendix.[1]

Data Simulation. We run 1,000 simulations. Each simulation involves generating returns for $N = 100$ stocks for $T = 1{,}150$ trading periods. Each trading period, the returns of all N stocks are governed by the returns of a subset of $K = 5$ stocks, $\mathcal{K}_t$, together with an idiosyncratic shock,

$$r_{n,t} = 0.15 \cdot \textstyle\sum_{\mathcal{K}_t} r_{k,t-1} + 0.001 \cdot \epsilon_{n,t}, \tag{17}$$

where $\epsilon_{n,t} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$. This collection of $K = 5$ sparse signals changes over time, leading to the time subscript on $\mathcal{K}_t$. We assume that there is a 1% chance that each signal

[1]See https://gist.github.com/alexchinco/467325abbf11d5c8f565.

λ Distribution, Simulated Data with Sparse Shocks

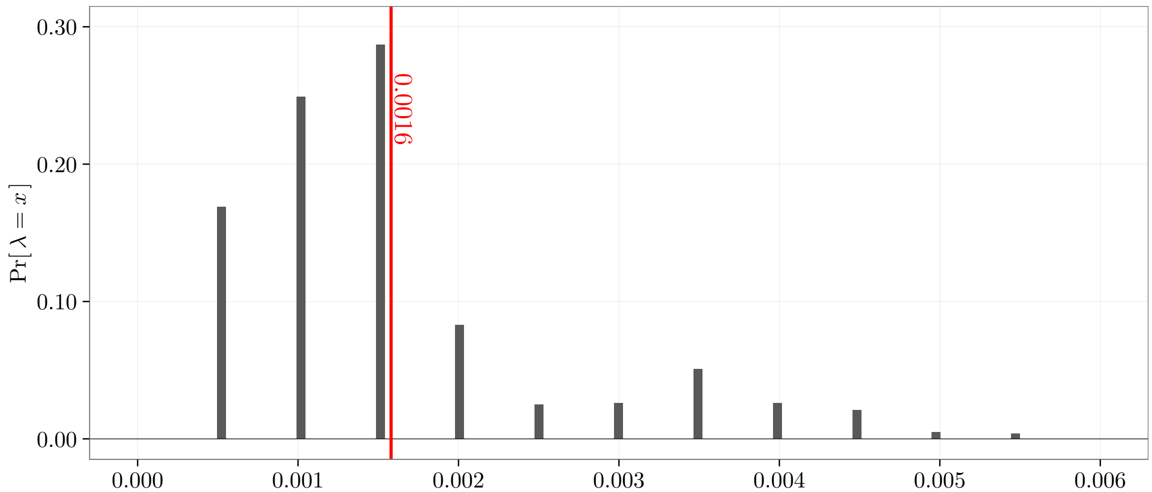

Figure 14: *Distribution of the optimal choice of penalty parameters, λ, estimated using the first* 150 *trading periods in the simulated data generated from Equation* (17). *Black bars: Probability that the optimal penalty parameter falls within a* 0.0005 *interval. Red vertical line: Average choice of penalty parameters. The discrete* 0.0005 *jumps come from the discrete grid of possible λs that we considered when running the code. Reads: "The estimation procedure typically picks out a penalty parameter of $\lambda = 0.0016$ in the simulated data."*

changes every period, so signals last $^{(1-0.01)}/_{0.01} = 99$ trading periods on average.

Fitting the Model. For each trading period from $t = 151$ to $t = 1{,}150$, we estimate the LASSO on the first stock, $n = 1$, using the previous $T = 50$ periods of data where the N possible predictors are the $N = 100$ stocks. This means using $T = 50$ time periods to estimate a model with $N = 100$ potential right-hand-side variables. As a useful benchmark, we also estimate the OLS model and an oracle model. In the oracle specification, we estimate an OLS regression with the $K = 5$ true predictors as the right-hand-side variables. Obviously, in the real-world you don't know what the true predictors are, but this specification gives an estimate of the best fit you could possibly achieve if you knew exactly where to look. After fitting each model to the previous 50 periods of data, we then make an out-of-sample forecast in the 51st period. The procedure is exactly the same as in Section 3.

Forecasting Regressions. We then check how closely these forecasts line up with the realized returns of asset $n = 1$ by analyzing the adjusted R^2 statistics from many forecasting regressions. For example, we take the LASSO's return forecast in trading

Predictor Distribution, Simulated Data with Sparse Shocks

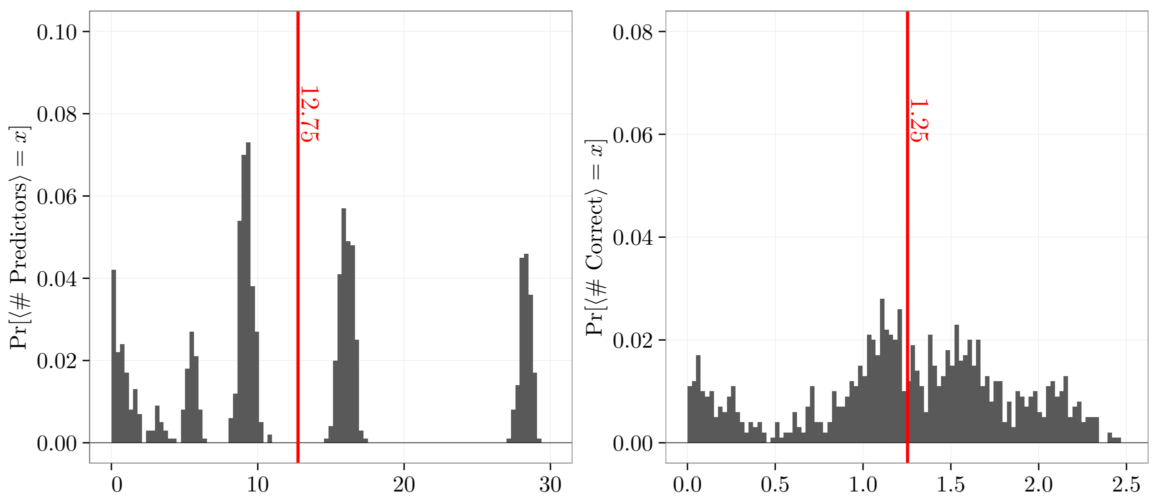

Figure 15: *Distribution of the number of predictors used by the LASSO when making its return forecast using simulated data generated from Equation (17). Left panel, black bars: Probability that the number of predictors falls within a* 0.5 *interval. Left panel, red vertical line: Average number of predictors used by the LASSO to make its return forecast. Right panel, black bars: Probability that the number of correct predictors chosen by the LASSO to make its return forecast falls within a* 0.05 *interval. Left panel, red vertical line: Average number of correct predictors chosen by the LASSO. Reads: "The LASSO usually only picks out the most important of the* $K = 5$ *correct predictors."*

periods $t = 151$ to $t = 1{,}150$ and estimate the regression below,

$$r_{1,t+1} = \tilde{a}_1 + \tilde{b}_1 \times \left(\frac{f_{1,t}^{\text{LASSO}} - \mu_1^{\text{LASSO}}}{\sigma_1^{\text{LASSO}}} \right) + \varepsilon_{1,t+1}, \tag{18}$$

where $\tilde{a}_1$ and $\tilde{b}_1$ are estimated coefficients, $r_{1,t+1}$ denotes the first stock's realized return in period $(t+1)$, $f_{1,t}^{\text{LASSO}}$ denotes the LASSO's forecast of the first stock's return in minute $(t+1)$, μ_1^{LASSO} and σ_1^{LASSO} represent the mean and standard deviation of this out-of-sample return forecast from period $t = 151$ to $t = 1{,}150$, and $\varepsilon_{1,t+1}$ is the regression residual. Figure 13 shows that the average adjusted-R^2 statistic from these 1,000 simulations is 4.40% for the LASSO; whereas, this statistic is only 1.29% when making your return forecasts using an OLS model.

Penalty Parameter Choice. Fitting the LASSO to the data involves selecting a penalty parameter, λ. We do this by selecting the penalty parameter that has the highest out-of-sample forecasting R^2 (equivalently Akaike information criterion (AIC)) during the first 100 periods of the data. This is why the forecasting regressions

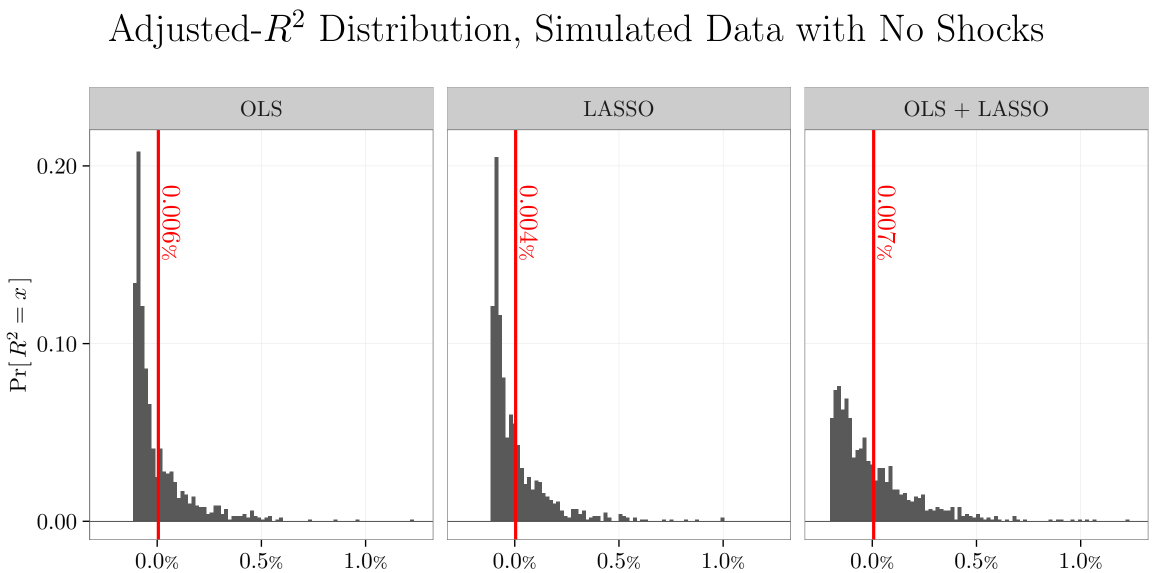

Figure 16: *Distribution of adjusted R^2s from the forecasting regressions using simulated data generated from Equation (19) where there are no shocks. Black bars: Probability that the adjusted R^2 from a single out-of-sample forecasting regression falls within a 0.1%-point interval. Red vertical line: Average adjusted R^2 from these regressions. Left panel: Out-of-sample prediction made using OLS. Center panel: Out-of-sample prediction made using the LASSO. Right panel: Out-of-sample predictions made using both OLS and the LASSO. Reads: "When there are no shocks, the LASSO does not add any forecasting power."*

above only use data starting at $t = 151$ instead of $t = 51$. Figure 14 shows the distribution of penalty parameter choices across the 1,000 simulations. The discrete 0.0005 jumps come from the discrete grid of possible λs that we considered when running the code.

Number of Predictors. If you look at the panel labeled "Oracle" in the adjusted R^2 figure, you'll notice that the LASSO's out-of-sample forecasting power is about a third of the true model's forecasting power, $^{4.40}/_{12.84} = 0.34$. This is because the LASSO doesn't do a perfect job of picking out the $K = 5$ sparse signals. The right panel of the figure below shows that the LASSO usually only picks out the most important of these $K = 5$ signals. What's more, the left panel shows that the LASSO also locks onto lots of spurious signals, suggesting that you might be able to improve the LASSO's forecasting power by choosing a higher penalty parameter, λ.

Placebo Tests. We conclude this section by looking at two alternative simulations where the LASSO shouldn't add any forecasting power. In the first alternative setting, Adjusted-R^2 Distribution, Simulated Data with Dense Shocks

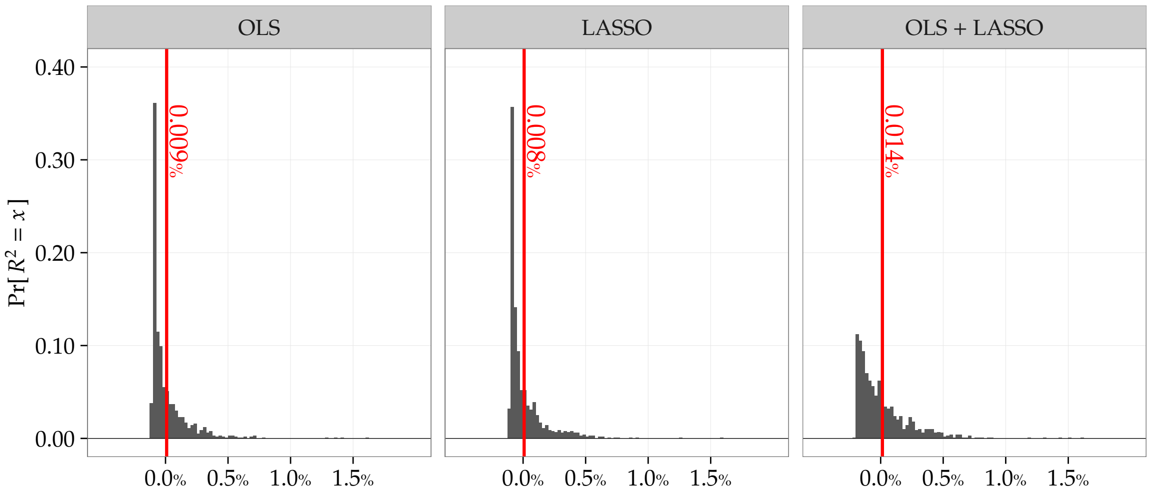

Figure 17: *Distribution of adjusted R^2s from the forecasting regressions using simulated data generated from Equation (17), but where there are $K = 75$ rather than $K = 5$ shocks. Black bars: Probability that the adjusted R^2 from a single out-of-sample forecasting regression falls within a 0.1%-point interval. Red vertical line: Average adjusted R^2 from these regressions. Left panel: Out-of-sample prediction made using OLS. Center panel: Out-of-sample prediction made using the LASSO. Right panel: Out-of-sample predictions made using both OLS and the LASSO. Reads: "When shocks are dense, the LASSO does not add any forecasting power."*

there are no shocks. That is, the returns for the $N = 100$ stocks are simulated using the model below,

$$r_{n,t} = 0.00 \cdot \textstyle\sum_{\mathcal{K}_t} r_{k,t-1} + \sigma \cdot \epsilon_{n,t}. \tag{19}$$

In the second setting, there are too many shocks: $K = 75$. Figure 16 and 17 show that, in both these settings, the LASSO doesn't add any forecasting power. Thus, running these simulations offers a pair of nice placebo tests showing that the LASSO really is picking up sparse signals in the cross-section of returns.

B Returns to Sparse Inference

The fact that using the LASSO boosts out-of-sample return predictability by quickly identifying the right predictors at the right time suggests that traders can use sparse-inference strategies, such as the LASSO, to quickly identify potential investment

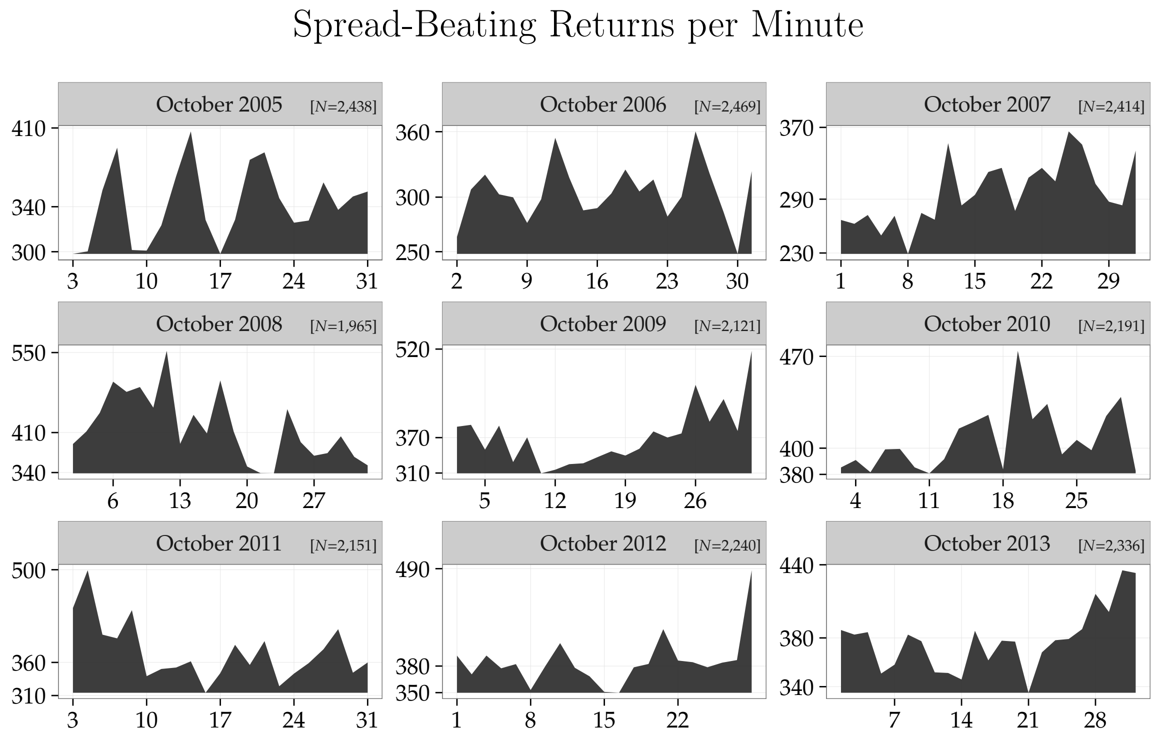

Figure 18: *Daily average number of stocks with spread-beating returns, $|r_{n,t+1}| > spread_{n,t}$ each minute. Data: NYSE-listed stocks traded in each October from 2005 to 2013. The spread is the average national-best bid-offer (NBBO) spread for each stock in a given minute. Days labeled on the x-axis are Mondays. The values labeled on the y-axis correspond to the minimum, median, and maximum of the number of spread-beating returns per minute during the month. Number in square brackets denote the number of NYSE-listed stocks each month. Reads: "Out of the* 2,121 *NYSE-listed stocks in our sample during October 2009, only* 310 *realized returns in excess of their NBBO spread on Friday, October 9th each minute."*

opportunities. This section directly verifies this observation by computing the returns to a trading strategy that buys or sells a stock whenever the LASSO's return forecast exceeds the bid-ask spread. This plain-vanilla trading strategy generates returns of 0.41% per month net of trading costs.

B.1 Trading-Strategy Description

The first column of Table 5 estimates a similar predictive regression as in Equation (8), but with only the LASSO's normalized prediction on the right-hand side,

$$r_{n,t+1} = \tilde{a}_n + \tilde{c}_n \cdot \left(\frac{f_{n,t}^{\text{LASSO}} - \tilde{\mu}_n^{\text{LASSO}}}{\tilde{\sigma}_n^{\text{LASSO}}} \right) + e_{n,t+1}. \tag{20}$$

We find that, for a typical stock, the average return to a market-timing strategy which is long when the LASSO's prediction is higher than average and short otherwise (Moskowitz et al., 2012) is $(390 \cdot 21) \cdot (3.18\times 10^{-4}) = 2.60\%$ per month using the same logic as described in Equation (5) from Subsection 2.1. But, this interpretation is subject to a pair of implementation-related caveats: it suffers from look-ahead bias and it ignores trading costs. We now analyze the returns to a trading strategy that corrects for these concerns.

Look-Ahead Bias. First, let's consider the problem of look-ahead bias. The issue is that when we computed the mean and standard deviation of our LASSO return forecast in Equation (20), we used information from future trading periods. For instance, the strategy dictated by Equation (20) is using information from October 26th, 2009 when deciding how many shares to buy on October 1st, 2009. To get around this problem, we split our sample in half each month and use the first 10 trading days of each October—that is, minutes $t = 1$ through $t = (293 \times 10) = 2{,}930$—to compute the mean and volatility of the out-of-sample LASSO predictions for each stock,

$$\begin{aligned} \hat{\mu}_n^{\text{LASSO}} &= \tfrac{1}{2{,}930} \cdot \textstyle\sum_{t=1}^{2{,}930} f_{n,t}^{\text{LASSO}} \\ \text{and} \quad \hat{\sigma}_n^{\text{LASSO}} &= \left(\tfrac{1}{2{,}930} \cdot \textstyle\sum_{t=1}^{2{,}930} \left[\, f_{n,t}^{\text{LASSO}} - \hat{\mu}_n^{\text{LASSO}} \,\right]^2 \right)^{1/2}. \end{aligned}$$

We then compute the returns to a trading strategy that is long whenever the prediction is positive and short whenever the prediction is negative,

$$r_{n,t+1}^{\text{LASSO}} = \left(\tfrac{f_{n,t}^{\text{LASSO}} - 0}{\hat{\sigma}_n^{\text{LASSO}}} \right) \times r_{n,t+1},$$

for the second half of each October from 2005 to 2013–that is, day 11 through the end of the trading month. By estimating each predictor's volatility in an earlier period and assuming each predictor's mean is zero, we avoid the look-ahead bias. All the information we need to compute the portfolio weights is available prior to the start of trading each minute.

This trading strategy buys a stock whenever the LASSO's out-of-sample return forecast is positive, $f_{n,t}^{\text{LASSO}} > 0$, and sells a stock whenever the LASSO's out-of-sample return forecast is negative, $f_{n,t}^{\text{LASSO}} < 0$. Moreover, for a given prediction, the strategy dictates that we trade more in stocks where the LASSO's out-of-sample return forecast is less volatile. We choose this portfolio weighting scheme in order to mirror the coefficients in the predictive regressions, not because it is somehow the optimal way to trade. The goal of this analysis isn't just to show that you can make money using the LASSO. Rather, we study the returns to a LASSO-based trading strategy because they provide evidence that the sparse signals we identify using the LASSO are economically important, that the sparse signals matter to real-world traders.

Trading Costs. Let's now turn our attention to the second problem—namely, trading costs—which are substantial when trading every minute. Figure 18 highlights this basic point by showing that, out of the roughly 2,000 NYSE-listed stocks in our sample each October, only around 364 realize returns in excess of their national-best bid-offer spread in any given minute. A predictor can be very good at forecasting small return fluctuations but be unhelpful because it doesn't predict the $^{364}/_{2{,}000} \approx 18\%$ of stocks each minute with price movements large enough to trade on.

We account for trading costs by redefining the strategy so that it only trades when the LASSO's out-of-sample return forecast exceeds the national-best bid-offer (NBBO) spread:

$$r_{n,t+1}^{\text{LASSO}} = \left\{ \left| \tfrac{f_{n,t}^{\text{LASSO}} - 0}{\hat{\sigma}_n^{\text{LASSO}}} \right| \times \left(\text{sgn}[f_{n,t}^{\text{LASSO}}] \cdot r_{n,t+1} - \mathit{spread}_{n,t} \right) \right\} \cdot 1_{\{|f_{n,t}^{\text{LASSO}}| > \mathit{spread}_{n,t}\}}.$$

So, for example, if $\mathit{spread}_{n,t} = 0$, then there is no spread and the strategy is the same as before. By contrast, if the spread is positive, $\mathit{spread}_{n,t} > 0$, then the trading

Implementation of Predictive Regressions

	Predictive Regression	Trading Strategy No Spread	Trading Strategy NBBO Spread
$\langle \tilde{a}_n \rangle_{\times 10^4}$	0.01 (0.92)		
$\langle \tilde{c}_n \rangle_{\times 10^4}$	3.18 (12.06)	3.46 (8.47)	0.41 (2.06)

Table 5: *Comparison of the monthly returns to a LASSO-based trading strategy as implied by the predictive regression in Equation (20) (labeled: predictive regression, column 1) and the realized monthly returns to a LASSO-based trading strategy (labeled: trading strategy, columns 2 and 3). In column 2, strategy buys or sells a stock whenever the normalized LASSO's return forecast is positive or negative, and no spread is paid. In column 3: strategy buys or sells a stock whenever the LASSO's return forecast exceeds the national best bid-offer (NBBO) spread, and the resulting return calculations include the cost of paying the spread. Numbers in parentheses are t-statistics and are clustered by (permno, month). Reads: "The trading strategy based on LASSO return forecast generates a* 0.41% *per month return net of trading costs when applied to an average stock."*

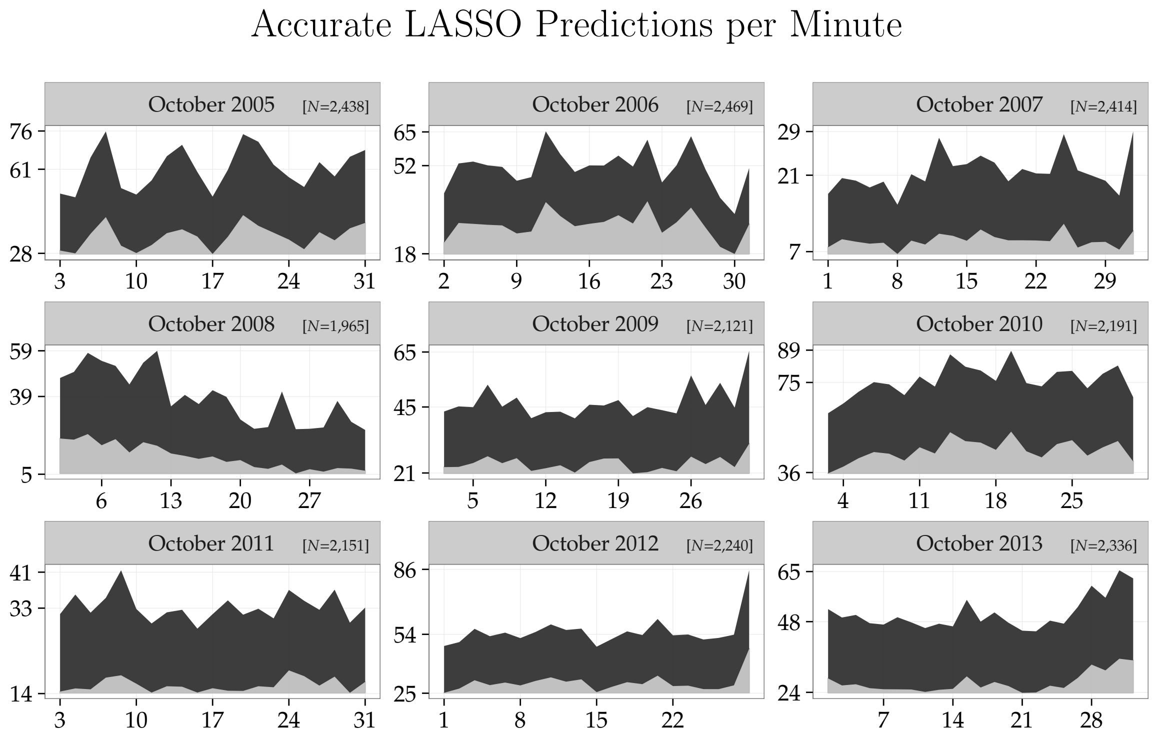

Figure 19: *Number of return forecasts made by the LASSO that end up beating the spread. Dark shaded: Daily average number of stocks each minute with spread-beating returns,* $|r_{n,t+1}| > spread_{n,t}$*, that were accurately predicted by only the LASSO,* $|f_{n,t}^{LASSO}| > spread_{n,t}$*. Light shaded: Daily average number of stocks each minute with spread-beating returns that were accurately predicted by both OLS and the LASSO. Days labeled on the* x*-axis are Mondays. The values labeled on the* y*-axis correspond to the minimum, median, and maximum of the number of accurate prediction made by the LASSO each the month. Data: NYSE-listed stocks traded in each October from 2005 to 2013. Reads: "Less than half of the LASSO's accurate predictions can be captured using an OLS regression."*

strategy only invests when the LASSO's return forecast is sufficiently large, $|f_{n,t}^{\text{LASSO}}| > spread_{n,t}$. Moreover, for a trade to be profitable, the LASSO's return forecast has to have both the right sign as the realized return in the next minute, $\text{sgn}[f_{n,t}^{\text{LASSO}}] \cdot r_{n,t+1} > 0$, and the realized return has to exceed the spread, $|r_{n,t+1}| > spread_{n,t}$. For each stock, we define the spread as the time-weighted average of the NBBO spread each minute in the TAQ data.

B.2 Realized Returns

We now investigate the returns to this LASSO-based trading strategy.

Estimation Results. Table 5 describes the returns per minute to trading strategies based on the LASSO's return forecasts under two different regimes: no spread and NBBO spread. For a typical NYSE-listed stock, the LASSO-based strategy generate positive gross returns of 3.46% per month in the absence of any trading costs. This point estimate is very close to the $(390 \cdot 21) \cdot (3.18 \times 10^{-4}) = 2.60\%$ per month point estimate we got when ignoring look-ahead bias in the second column of Table 1.

As you would expect, introducing trading costs dramatically lowers the trading strategy's returns. After accounting for the spread, the LASSO-based trading strategy has a net return of 0.41% per month when applied to a typical stock. But, this return is still positive and statistically significant. You could engineer a more sophisticated LASSO-based strategy to deliver much larger returns, but that isn't our goal here. These positive returns are interesting because they show that the sparse signals that the LASSO is using to make its return forecasts are economically important.

Accurate Predictions. Figure 19 shows the daily average number of accurate predictions made by the LASSO each trading day. These are stock-minutes where the stock realized a return in excess of its NBBO spread, $|r_{n,t+1}| > \mathit{spread}_{n,t}$, and where the LASSO said the stock would realize a spread-beating return, $|f^{\text{LASSO}}_{n,t}| > \mathit{spread}_{n,t}$. The dark regions represent the number of accurate predictions that were only made by the LASSO. The lighter regions represent the number of accurate predictions that were made by both the LASSO and an OLS regression. The LASSO typically picks out only around 60 of the 364 possible spread-beating returns each minute.

Different Predictions. What's more, less than half of the LASSO's 60 accurate predictions can be captured using an OLS regression. For example, the probability that both OLS and the LASSO select the same stock to beat the spread in a given minute is 6%. Each strategy generates a different pattern of returns because each strategy tells traders to hold very different collections of assets. This is another way of showing that the LASSO and OLS are capturing very different kinds of information.

C Additional Results

Out-of-Sample Return Predictability, Other AR(L) Choices

	AR(1)		AR(2)		AR(4)		AR(5)	
$\langle \tilde{a}_n \rangle \times 10^4$	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)
$\langle \tilde{b}_n \rangle \times 10^4$	4.10 (13.68)	3.46 (14.46)	4.44 (13.71)	3.86 (14.54)	4.37 (13.71)	3.81 (14.74)	4.21 (13.56)	3.65 (14.66)
$\langle \tilde{c}_n \rangle \times 10^4$		2.06 (12.68)		1.92 (13.19)		2.04 (13.20)		2.15 (13.16)
$\langle \text{Adj. } R^2 \rangle$	7.07%	9.03%	8.21%	10.00%	7.88%	9.90%	7.33%	9.53%
$\langle \Delta_{\text{Adj. } R^2} \rangle$	1.95 (7.61)		1.78 (8.40)		2.02 (8.91)		2.20 (8.89)	
$\langle F\text{-test} \rangle$	10.77		10.33		11.08		11.57	

Table 6: *Average of the parameter estimates from the out-of-sample regressions each month described by Equations (4) and (8) with alternative specifications for the benchmark OLS model. Panel a): benchmark model for stock n includes 3 lagged returns for stock n and the market return. Panel b): benchmark model for stock n includes 3 lagged returns for stock n and 3 lagged returns of the stock with the highest pairwise correlation with stock n in the previous trading day. Panel c): benchmark model for stock n includes the optimal number of lagged returns for stock n according to the Akaike information criterion. Sample period: (stock,minute)-level regressions in each October from 2005 to 2013. Coefficient estimates have units of percent per minute. Numbers in parentheses are the t-statistics and are clustered by (permno, month). Δ_n is the difference between the adjusted R^2 when using both the LASSO and OLS return forecasts and the adjusted R^2 when using only the OLS return forecast. F-test, p-value reports the probability that the LASSO's return forecast adds out-of-sample predictive power.*

Out-of-Sample Return Predictability, Benchmark with Market Returns vs. Benchmark with Noise

	AR(3)	AR(3) + r_{Mkt}	AR(3) + N(0,1)
$\langle \tilde{a}_n \rangle_{\times 10^4}$	0.01 (0.92)	0.01 (0.92)	0.01 (0.92)
$\langle \tilde{b}_n \rangle_{\times 10^4}$	4.45 (13.76)	3.98 (13.49)	4.00 (13.52)
$\langle$Adj. $R^2\rangle$	8.17%	6.65%	6.71%

Table 7: *Average of the parameter estimates from the out-of-sample regressions each month described by Equation (4) with alternative specifications for the benchmark OLS model.* AR(3)*: benchmark model for stock n includes* 3 *lagged returns for stock n.* AR(3) + r_{Mkt}*: benchmark model for stock n includes* 3 *lagged returns for stock n and* 3 *lags of the market return.* AR(3) + N(0, 1)*: benchmark model for stock n includes* 3 *lagged returns for stock n and* 3 *iid draws from a standard normal distribution. Sample period: (stock,minute)-level regressions in each October from 2005 to 2013. Coefficient estimates have units of percent per minute. Numbers in parentheses are the t-statistics and are clustered by (permno, month). Reads: "Adding more variables to the benchmark model will clearly raise the R^2 of this model in sample, but an additional superfluous variable will lower the model's adjusted R^2 out-of-sample due to overfitting bias."*