

# Sparse Signals in the Cross-Section of Returns\*

Alex Chinco,<sup>†</sup> Adam D. Clark-Joseph,<sup>‡</sup> and Mao Ye<sup>§</sup>

December 31, 2015

## Abstract

This paper applies the least absolute shrinkage and selection operator (LASSO) to identify rare, short-lived, “sparse” signals in the cross-section of returns. The LASSO is an ordinary-least-squares (OLS) regression combined with a penalty function that shrinks small OLS coefficients to be exactly zero, so it is well-defined even when there are many more predictors than observations. Using the LASSO increases out-of-sample predictability in minute-by-minute NYSE returns by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ , and generates trading-strategy returns of 0.30% per month net of trading costs. This predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some persistent factor. The LASSO typically forecasts a stock’s returns using the lags of only 11 other stocks (a mere 0.5% of all possible choices), and 90% of these predictors last 4 minutes or less. This success implies that returns have a sparse structure and suggests a new way of thinking about the economic forces behind returns.

JEL CLASSIFICATION. C55, C58, G12, G14

KEYWORDS. The LASSO, Sparsity, High-Dimensional Inference, Return Predictability

---

\*We have received many helpful comments and suggestions from John Campbell, Xavier Gabaix, Andrew Karolyi, Maureen O’Hara, Thomas Ruchti, Gideon Saar, Heather Tookes, Sunil Wahal, and Brian Weller as well as seminar participants at the University of Illinois Urbana-Champaign and the 11th Annual Central Bank Conference on the Microstructure of Financial Markets. This research is supported by National Science Foundation grant 1352936. 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>

<sup>†</sup>University of Illinois at Urbana-Champaign, [alexchinco@gmail.com](mailto:alexchinco@gmail.com), (916) 709-9934.

<sup>‡</sup>University of Illinois at Urbana-Champaign, [adcj@illinois.edu](mailto:adcj@illinois.edu), (217) 244-1536.

<sup>§</sup>University of Illinois at Urbana-Champaign, [maoye@illinois.edu](mailto:maoye@illinois.edu), (217) 244-0474.

# 1 Introduction

Financial economists have been looking for variables that forecast future stock returns for as long as there have been financial economists. For some recent examples, think about [Jegadeesh and Titman \(1993\)](#), who show that a stock's current returns are predicted by the stock's returns over the previous 12 months, [Hou \(2007\)](#), who shows that the current returns of smallest stocks in an industry are predicted by the lagged returns of the largest stocks in the industry, and [Cohen and Frazzini \(2008\)](#), who show that a stock's current returns are predicted by the lagged returns of its major customers.

When you think about it, finding these sorts of forecasting variables actually consists of two separate problems, identification and estimation. Until now these two separate problems have always been tackled using two separate toolkits. First, researchers have used their intuition to identify a new predictor—let's call it  $x_t$ . Then, they have used statistics to estimate this new predictor's quality,

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

where  $\hat{\theta}_0$  and  $\hat{\theta}_1$  are estimated coefficients,  $r_{n,t+1}$  is the return on the  $n$ th stock, and  $\epsilon_{n,t+1}$  is the regression residual. If knowing  $x_t$  reveals a lot of information about what a stock's future returns will be, then  $|\hat{\theta}_1|$  and the associated  $R^2$  will be large.

But, modern financial markets are big, fast, and densely interconnected. Predictability doesn't always occur at scales that are easy for researchers to intuit, making the standard approach to tackling the first problem problematic. For instance, the lagged returns of the Family Dollar Corporation were a significant predictor for more than 25% of all NYSE-listed oil and gas stocks during a 20-minute stretch on October 6th, 2010. Can a researcher really fish this particular forecasting variable out of a sea of spurious predictors using intuition alone? And, how exactly is he supposed to do this in under 34 minutes?

This paper replaces intuition with statistics and uses the least absolute shrinkage and selection operator (LASSO) to identify rare, short-lived, "sparse" signals in the cross-section of returns. The LASSO increases out-of-sample predictability in minute-by-minute NYSE returns by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ , and thereby suggests a new way of thinking about the economic forces behind stock returns.

*Estimation Strategy.* We begin our analysis by asking: What's so special about the LASSO? Why not just add in the lagged returns of other NYSE-listed stocks when estimating Equation (1)? This is a natural first thought, but the problem with this approach is that it implicitly assumes all possible cross-stock relationships are equally important. For instance, there are 2,191 NYSE-listed stocks in our data for October 2010, so the resulting

estimation problem for this one month would have 2,192 free parameters: one for the intercept and one for the return of each NYSE stock in the previous minute. Estimating all of these free parameters using an ordinary least squares (OLS) regression would require at least 2,192 minutes of data, which is nearly 6 trading days! To shorten the required sample length, we need to take a different approach and focus on only the most important predictors.

The LASSO allows us to do just that. This penalized-regression technique, which was introduced in Tibshirani (1996), simultaneously identifies and estimates the most important coefficients using a far shorter sample period by betting on sparsity—that is, by assuming only a handful of variables actually matter at any point in time. Formally, using the LASSO means solving the problem below,

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

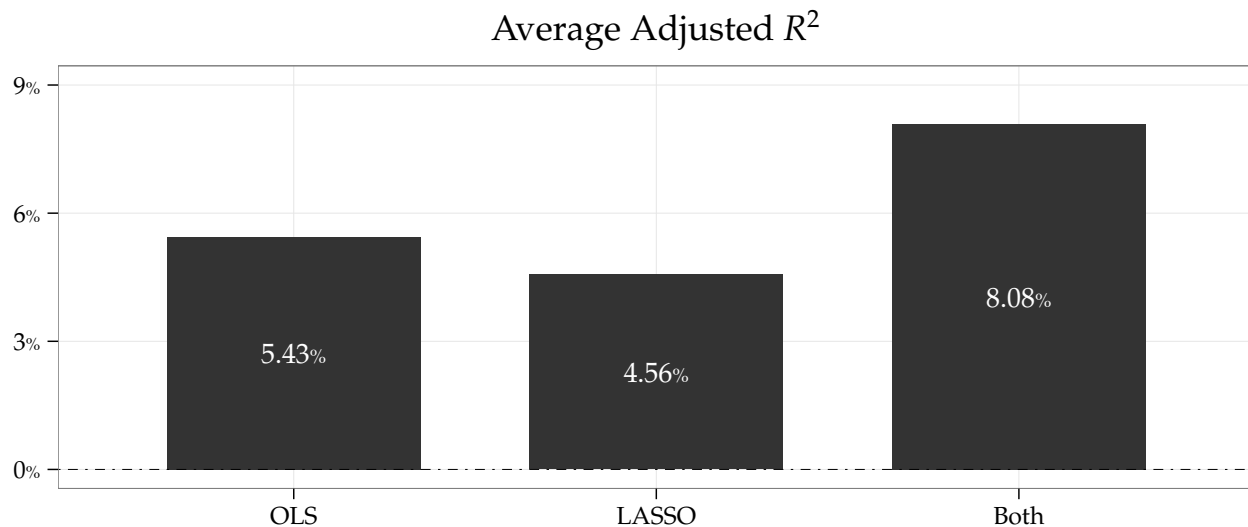
where  $r_{n,t}$  is the  $n$ th stock's return in minute  $t$ ,  $\hat{\vartheta}$  is a  $(2,192 \times 1)$ -dimensional vector of estimated coefficients,  $T$  is the number of minutes in the sample period, and  $\lambda$  is a penalty parameter. Equation (2) looks complicated at first, but it's not. It's 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's this penalty function that's 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 better understand how the LASSO does this, consider the solution to Equation (2) when the right-hand-side variables are uncorrelated and have unit variance:

$$\hat{\vartheta}_{n'} = \text{sgn}[\hat{\theta}_{n'}] \cdot (|\hat{\theta}_{n'}| - \lambda)_+. \quad (3)$$

Here,  $\hat{\theta}_{n'}$  represents what the standard OLS coefficient would have been if we had an infinite amount of data,  $\text{sgn}[x] = x/|x|$ , and  $(x)_+ = \max\{0, x\}$ . On one hand, this solution means that, 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, the solution implies that, if OLS would have estimated a sufficiently small coefficient,  $|\hat{\theta}_{n'}| < \lambda$ , then the LASSO is going to pick  $\hat{\vartheta}_{n'} = 0$ . 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 possible predictors,  $T \ll 2,192$ . 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 select and then estimate the size of these few important coefficients.

*Out-of-Sample Predictability.* As a benchmark, we first use OLS to estimate rolling



**Figure 1:** Average adjusted  $R^2$  from out-of-sample forecasting regressions using the return-forecasts from an OLS regression, the LASSO, or both. Data: Minute-by-minute returns for NYSE-listed stocks in each October from 2005 to 2013. Reads: “Including the LASSO’s return-forecast boosts out-of-sample return predictability by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ .”

autoregressions with 30 minutes of data and find that the average out-of-sample adjusted  $R^2$  is 5.43% for NYSE-listed stocks. That is, on average 5.43% 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. Having estimated this benchmark model, we next consider the effects of other stocks’ returns over the previous 3 minutes. This means using 30 minutes of data to both identify and estimate the handful of significant predictors from among  $1 + (3 \times 2,191) = 6,574$  possibilities in October 2010. Our main result is that including the LASSO’s return-forecast boosts the out-of-sample adjusted  $R^2$  by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ ! This result isn’t driven by a few outlying observations; rather, it’s a robust feature of all stocks over our entire sample period.

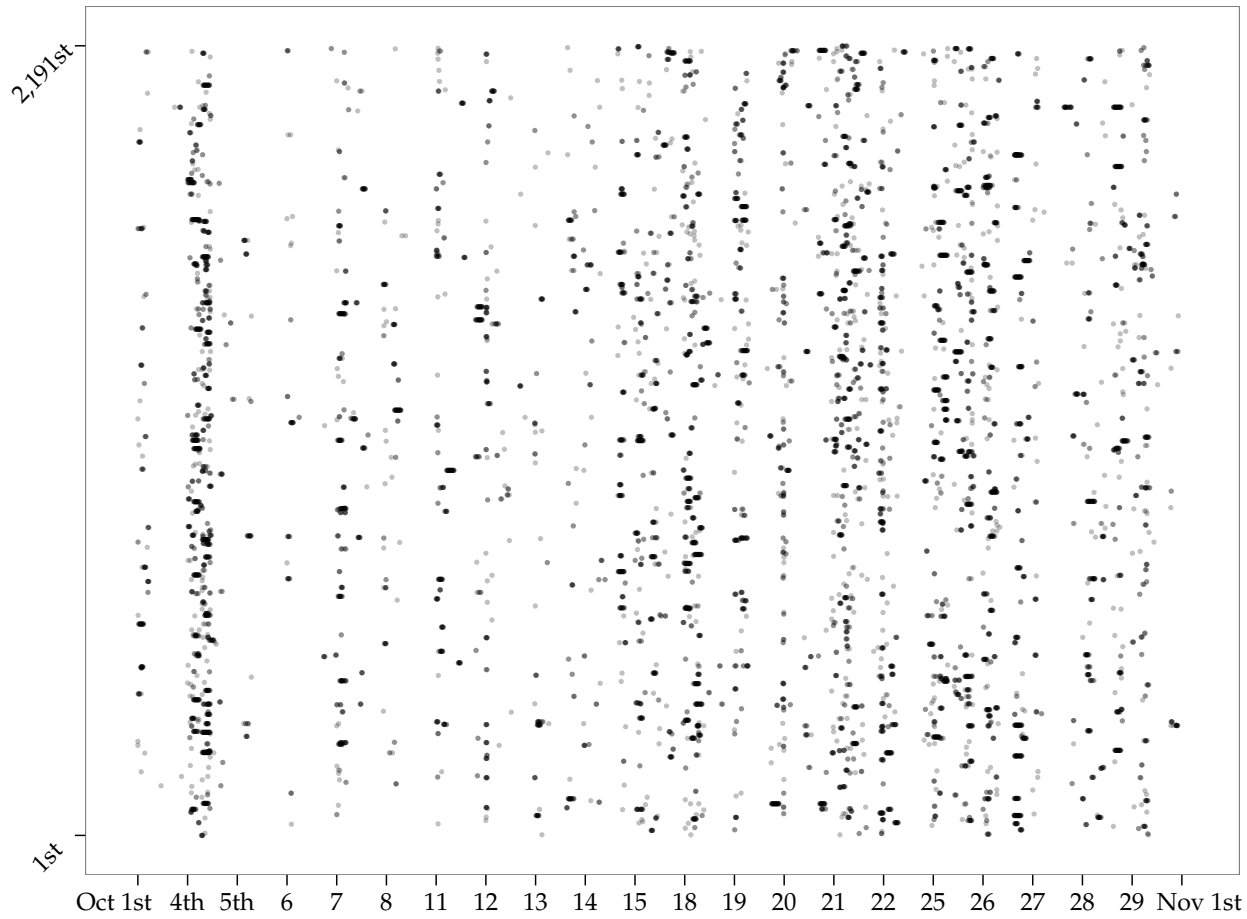
*Trading-Strategy Returns.* Next, to show that this predictability isn’t just a statistical artifact, we compute the returns to a trading strategy that buys or sells a stock whenever the LASSO’s return-forecast exceeds the bid-ask spread. This plain-vanilla strategy generates returns of 0.30% per month net of trading costs, and these positive net returns exist in each subsample of the data we look at. The goal of this analysis isn’t just to show that you can make money using the LASSO—after all, this is an academic paper. 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.

*Evidence of Sparsity.* After documenting the LASSO's predictive power and showing that this predictability matters to real-world traders, we next ask ourselves: what sort of information is the LASSO picking up? It turns out that the LASSO only selects around 11 predictors on average. That is, when making its return-forecast for a single stock, such as Exxon, the LASSO typically considers the lagged returns of only 11 other stocks during the previous 3 minutes. This is only around 0.5% of the roughly 2,000 stocks that it can choose from each month! What's more, the set of significant predictors changes rapidly. If the LASSO is using the lagged returns of the Family Dollar Corporation to predict Exxon's future returns right now, then there is only a 10% chance that the LASSO will still be using Family Dollar's lagged returns in 5 minutes. Finally, the LASSO tends to load on the same predictors when making forecasts for different stocks. 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 this variable when making return-forecasts for other stocks, such as British Petroleum or Chevron. Thus, the LASSO's predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some persistent factor.

*More Than Just News.* We use data from RavenPack to study how the LASSO's choice of predictors is related to news announcements. We find that, for example, even though Family Dollar is more likely to be chosen by the LASSO as a significant predictor for some other stock's returns in the minutes following a news announcement about Family Dollar, there is nothing in the announcements that predicts which stocks Family Dollar's returns will help forecast. Should we look at oil and gas stocks like Exxon? Industrials like Mitsubishi? Or, somewhere else? In short, the LASSO is doing more than just mimicking information that's in news announcements. It's identifying the way in which this information is propagating through the market.

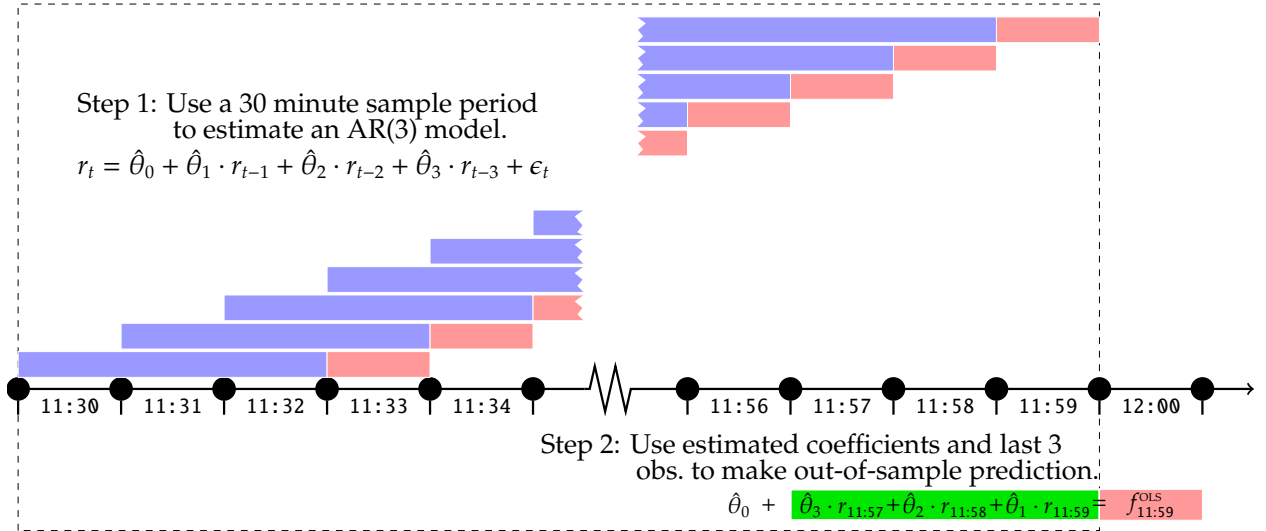
*Economic Implications.* Finally, we conclude our analysis by discussing the economic implications of the LASSO's success, which suggests a new way of thinking about the economic forces behind stock returns. If we only run OLS regressions like Equation (1), then it's hard to think about anything other than persistent factors,  $\hat{\theta}_1 \cdot x_t$ , and idiosyncratic noise,  $\epsilon_{n,t+1}$ , driving stock returns. But, we know that factors like market and industry returns only account for roughly 20% of asset-return volatility (see [Campbell, Lettau, Malkiel, and Xu \(2001\)](#)). So, why do returns move around so much? Can the remaining 80% really just be due to the accidents of life? This paper suggests an alternative. The LASSO's predictive power implies that returns realize sparse shocks. These sparse shocks are more structured than noise (see [Figure 2](#)) but can't be captured by OLS regressions.

## Significant Predictors in October 2010



**Figure 2:** Heat map showing the minutes during October 2010 during which the LASSO selected a given predictor when making its return-forecast for more than 2% of all NYSE-listed stocks. Darker dots indicate predictors that were used in more return-forecasts.

## Estimation Strategy and Timing



**Figure 3:** To make an out-of-sample predictions in minute  $(t + 1) = 12:00\text{pm}$  via ordinary least squares (OLS), we estimate an autoregressive model with  $L^*$  lags using the stock's returns in the previous 30 minutes, selecting the optimal number of lags using the Akaike information criterion. Then, we use the estimated coefficients to predict the stock's returns in minute  $(t + 1) = 12:00\text{pm}$ , referring to the prediction as  $E_t[r_{t+1}] = f_t^{\text{OLS}}$ . The figure above shows this process when the optimal number of lags is  $L^* = 3$ .

## 2 Out-of-Sample Predictability

We find that the minute-by-minute returns of NYSE-listed stocks are 1.5-times more predictable out-of-sample after using the LASSO to account for sparse signals. Our data consists of the minute-level returns of NYSE-listed stocks from TAQ during the month of October in each year from 2005 to 2013—that is, 9 years in total. We restrict the sample to stocks which had prices exceeding \$5 at the start of the month and which were traded every day of the month.

### 2.1 OLS Regression

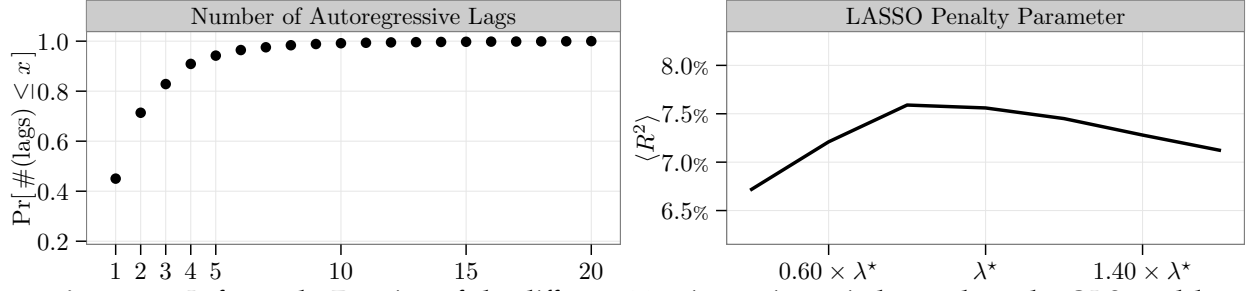
To provide a benchmark, we begin by making out-of-sample return-forecasts by fitting via ordinary least squares (OLS), an approach which explicitly does not take into account any sparse signals. Figure 3 outlines the timing of the estimation strategy.

*Estimation Strategy.* For each NYSE-listed stock in our data, we estimate a series of autoregressions in rolling 30-minute windows,

$$r_{n,t} = \hat{\theta}_0 + \sum_{\ell=1}^{L^*} \hat{\theta}_\ell \cdot r_{n,t-\ell} + \epsilon_{n,t}, \quad (4)$$

where  $\hat{\theta}_0$  and  $\{\hat{\theta}_\ell\}$  are estimated coefficients,  $L^*$  is the optimal number of lags for the stock

## Choice of Tuning Parameters



**Figure 4:** Left panel: Fraction of the different 30-minute time windows where the OLS model in Equation (5) 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%.”

during the 30-minute window according to the Akaike information criterion (AIC),  $r_{n,t}$  denotes stock the  $n$ th stock’s return in minute  $t$ , and  $\epsilon_{n,t}$  is the regression residual. The first 30-minute window we consider each day is  $t \in \{10:37\text{am}, 10:38\text{am}, \dots, 11:06\text{am}\}$  and the last window is  $t \in \{3:29\text{pm}, 3:30\text{pm}, \dots, 3:58\text{pm}\}$ , yielding 293 total such samples for each stock on each day. Our first prediction each trading day is at 11:07am and our last prediction each trading day is at 3:59pm. The Akaike information criterion chooses 3 or fewer lags in more than 82% of the 30-minute time windows that we study as shown in the left panel of Figure 4.

*Out-of-Sample Prediction.* Next, we predict each stock’s return in the 31st minute using the coefficient estimates from the preceding 30-minute training sample:

$$E_t[r_{n,t+1}] = f_{n,t}^{\text{OLS}} = \hat{\theta}_0 + \sum_{\ell=1}^{L^*} \hat{\theta}_\ell \cdot r_{n,t-\ell+1}. \quad (5)$$

So, for example, if we estimated the coefficients for IBM over the 30-minute window  $t \in \{11:30\text{am}, 11:31\text{am}, \dots, 11:59\text{am}\}$ , then we’d use these coefficients to predict IBM’s return in minute  $(t+1) = 12:00\text{pm}$ . This gives us 293 out-of-sample predictions for each stock each day, one for each 30-minute training period.

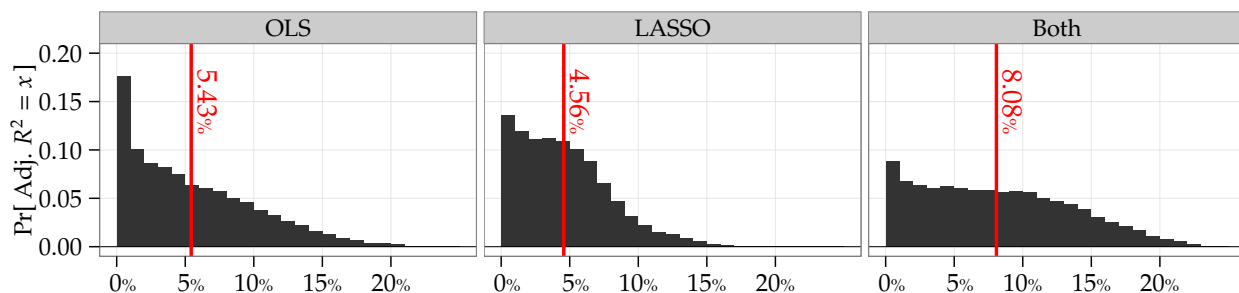
To test whether these out-of-sample predictions are good or bad, we regress the realized returns in the 31st minute on the normalized return-forecast for each stock,

$$r_{n,t+1} = \tilde{a}_n + \tilde{b}_n \cdot \left( \frac{f_{n,t}^{\text{OLS}} - \mu_n^{\text{OLS}}}{\sigma_n^{\text{OLS}}} \right) + e_{n,t+1}, \quad (6)$$

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 of stock  $n$ ’s return in minute  $(t+1)$  using an autoregression



## Distribution of Adjusted $R^2$ s



**Figure 5:** Distribution of adjusted  $R^2$ s from the forecasting regressions in Equations (6), (10), and (12). 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 (6). Middle panel: Out-of-sample prediction made using the LASSO as in Equation (10). Right panel: Out-of-sample predictions made using both OLS and the LASSO as in Equation (12). Reads: “While the OLS and LASSO models have similar out-of-sample fits on average, 5.43% vs. 4.56%, their fits display very different cross-sectional distributions meaning that each estimator is picking up very different information.”

comma

model,  $\mu_n^{\text{OLS}}$  and  $\sigma_n^{\text{OLS}}$  represent the mean and standard deviation of this out-of-sample prediction over the entire sample period, and  $e_{n,t+1}$  is the regression residual. To be clear, this means running separate regressions for each stock in each month—for example, one regression for each of the 2,192 NYSE-listed stocks in our sample in October 2010 and one regression for each of the 1,965 NYSE-listed stocks in our sample for October 2008.

*Estimation Results.* The first column of Table 1 shows that the average adjusted  $R^2$  in these regressions is 5.43%. That is, for a randomly selected stock, you can explain 5.43% of the variation in its minute-by-minute returns using only information about that stock’s past returns. Table 1 also shows this analysis for different subsamples of our data. This subsample analysis indicates that the average fit is relatively stable. If using the LASSO to account for sparse signals adds value, then it has to improve on this 5.43% benchmark.

While we are primarily interested in the model’s fit, it’s useful to look at the coefficients to gain some economic intuition about what the model is telling us. Specifically, we can interpret  $\tilde{b}_n$  as the average return per minute to a time-series momentum strategy à la Moskowitz, Ooi, and Pedersen (2012):

$$\tilde{b}_n = \frac{1}{T \cdot \sigma_n^{\text{OLS}}} \cdot \sum_{t=1}^T (f_{n,t}^{\text{OLS}} - \mu_n^{\text{OLS}}) \cdot r_{n,t+1}. \quad (7)$$

So, the average coefficient of  $\langle \tilde{b}_n \rangle = 3.57 \times 10^{-4}$  means that the gross monthly return to a market-timing strategy that is long a stock when the OLS model’s prediction is higher than average and short the stock otherwise is  $(390 \cdot 21) \cdot \langle \tilde{b}_n \rangle = 2.92\%$  per month.

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,  $\mu_n^{\text{OLS}}$  and  $\sigma_n^{\text{OLS}}$ , even though this information is not known at the beginning of the sample period. For these reasons, this 2.92% per month figure should be taken as an upper bound on the profitability of the OLS predictor, and we examine its usefulness as a trading predictor in the presence of transactions costs in Section 3 below.

## 2.2 Penalized Regression

Let's now consider the impact of using other stocks' returns over the previous 3 minutes using 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.

*Estimation Strategy.* For each of the NYSE-listed stocks in our sample each October, we compute a series of LASSO estimates using rolling 30-minute windows just like we did for the OLS-based approach. The LASSO solves the optimization problem below,

$$\hat{\vartheta} = \arg \min_{\vartheta \in \mathbb{R}^{7,636}} \left\{ \frac{1}{2 \cdot 30} \cdot \sum_{t=1}^{30} \left( r_{n,t} - \vartheta_0 - \sum_{n'=1}^N \sum_{\ell=1}^3 \vartheta_{n',\ell} \cdot r_{n',t-\ell} \right)^2 + \lambda \cdot \sum_{n'=1}^N \sum_{\ell=1}^3 |\vartheta_{n',\ell}| \right\}, \quad (8)$$

where  $\vartheta_{n,0}$  and  $\{\vartheta_{n',\ell}\}$  are estimated coefficients,  $r_{n,t-\ell}$  denotes stock the  $n$ th stock's return  $\ell$  minutes ago,  $\lambda$  is a penalty parameter, and  $N$  is the number of stocks in a given month. We perform this analysis on the exact same set of 30-minute training samples for each stock as in the subsection above. Our first prediction each trading day is at 11:07am and our last prediction each trading day is at 3:59pm. This gives us 293 out-of-sample predictions for each stock each day, one for each 30-minute training period. We give a detailed example of how the LASSO works using simulated data in Appendix A.

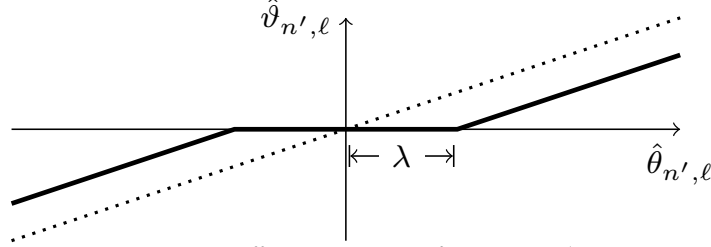
*Out-of-Sample Prediction.* To see whether or not the LASSO's coefficient estimates contain useful information, we create an out-of-sample prediction for each 30-minute training sample just as before,

$$E_t[r_{n,t+1}] = f_{n,t}^{\text{LASSO}} = \hat{\vartheta}_0 + \sum_{n'=1}^N \sum_{\ell=1}^3 \hat{\vartheta}_{n',\ell} \cdot r_{n',t-\ell+1}, \quad (9)$$

and then regress each stock's realized returns on the LASSO's normalized return-forecast,

$$r_{n,t+1} = \tilde{a}_n + \tilde{c}_n \cdot \left( \frac{f_{n,t}^{\text{LASSO}} - \mu_n^{\text{LASSO}}}{\sigma_n^{\text{LASSO}}} \right) + e_{n,t+1}, \quad (10)$$

## LASSO vs. OLS Estimates



**Figure 6:** *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  $\hat{\vartheta}_{n',\ell} = 0$ .”*

where  $\tilde{a}_n$  and  $\tilde{c}_n$  are estimated coefficients,  $r_{n,t+1}$  denotes the  $n$ th stock’s realized return in minute  $(t + 1)$ ,  $f_{n,t}^{\text{LASSO}}$  denotes our prediction of the  $n$ th stock’s return in minute  $(t + 1)$  using the LASSO,  $\mu_n^{\text{LASSO}}$  and  $\sigma_n^{\text{LASSO}}$  represent the mean and standard deviation of this out-of-sample prediction over the entire sample period, and  $e_{n,t+1}$  is the regression residual.

*Betting on Sparsity.* Before we discuss the results, let’s first ask ourselves: what additional information might the LASSO be capturing? To gain some intuition, consider the solution to the optimization problem in Equation (8) when the right-hand-side variables are uncorrelated and have unit variance,

$$\hat{\vartheta}_{n',\ell} = \text{sgn}[\hat{\theta}_{n',\ell}] \cdot (|\hat{\theta}_{n',\ell}| - \lambda)_+, \quad (11)$$

where  $\hat{\theta}_{n',\ell}$  represents what the OLS coefficient would have been given enough data,  $\text{sgn}[x] = x/|x|$ , and  $(x)_+ = \max\{0, x\}$ . Equation (11) says that, 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 6, you see that the solid line denoting the LASSO estimate and the dotted line denoting the OLS estimate are quite 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 Figure 6.

Thus, the LASSO’s return-forecast is helpful only under certain conditions. If there are only a few (that is,  $K \leq 30$ ) important predictors with coefficients larger than  $\lambda$  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. But, if there are no significant predictors or if these signals are not sparse (that is,  $K > 30$ ), then the LASSO’s return-forecast won’t be a helpful predictor. In the first case, there wouldn’t be any cross-stock signals to estimate. In the second case, there would be too many cross-stock signals to estimate using only 30 data points. It’s possible to bet on sparsity and lose.

*Estimation Results.* The second column of Table 1 displays summary statistics describ-

ing the results of the LASSO forecasting regressions. Looking at the average adjusted  $R^2$  from these regressions, we see that the fit of the LASSO's prediction is on par with that of the prediction from the autoregressive model. When we include the out-of-sample predictions from both the autoregression and the LASSO in the same regression,

$$r_{n,t+1} = \tilde{a}_n + \tilde{b}_n \cdot \left( \frac{f_{n,t}^{\text{OLS}} - \mu_n^{\text{OLS}}}{\sigma_n^{\text{OLS}}} \right) + \tilde{c}_n \cdot \left( \frac{f_{n,t}^{\text{LASSO}} - \mu_n^{\text{LASSO}}}{\sigma_n^{\text{LASSO}}} \right) + e_{n,t+1}, \quad (12)$$

we find that these two predictions are capturing very different kinds of information. The third column of Table 1, which displays the summary statistics from the combined regressions, reveals that including information from the LASSO increases the out-of-sample adjusted  $R^2$  for the typical stock by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ . Although the OLS model and the LASSO generate predictions with similar accuracies on average, each model uses very different information to make its forecast as shown in Figure 5. The bet on sparsity pays off.

comma

## 2.3 Robustness Checks

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

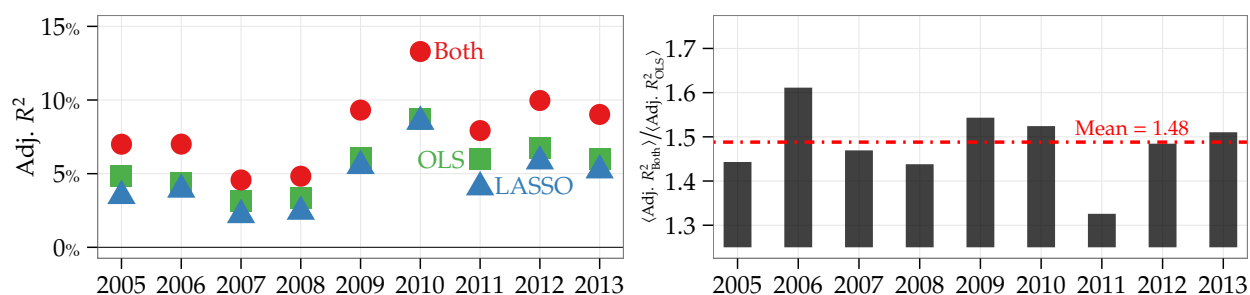
*Sample Periods.* Table 1 shows the summary statistics for the predictive regressions broken down into a set of 3 different 3-year intervals. We now further slice our results into year-specific segments to show that the gain from including the LASSO's return-forecast is consistent over time. The left panel of Figure 7 shows the average adjusted  $R^2$  from

Out-of-Sample Return Predictability

	Full Sample			2005-07		2008-10		2011-13	
$\langle \tilde{a}_n \rangle \times 10^{-4}$	0.01 (19.42)	0.01 (19.42)	0.01 (19.41)	0.01 (18.75)	0.01 (18.75)	0.00 (4.78)	0.00 (4.78)	0.02 (29.77)	0.02 (29.77)
$\langle \tilde{b}_n \rangle \times 10^{-4}$	3.57 (140.59)		3.00 (136.06)	2.71 (77.08)	2.35 (75.02)	4.16 (97.17)	3.45 (91.57)	3.97 (77.08)	3.28 (74.75)
$\langle \tilde{c}_n \rangle \times 10^{-4}$		3.17 (166.77)	2.40 (175.02)		1.91 (97.22)		2.76 (133.89)		2.59 (90.24)
$\langle \text{Adj. } R^2 \rangle$	5.43%	4.56%	8.08%	4.11%	6.20%	6.13%	9.30%	6.22%	8.98%

**Table 1:** Average of the parameter estimates from the out-of-sample regressions each month described by Equations (6), (10), and (12). The 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. Reads: "While the typical fit of the LASSO's prediction is on par with the typical fit of the OLS model's prediction, adjusted  $R^2 = 5.43\%$  vs. adjusted  $R^2 = 4.56\%$ , the two forecasts are using fundamentally different information. Including both the OLS and the LASSO forecasts increases out-of-sample return predictability by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ ."

## Distribution of Adjusted $R^2$ s, by Year



**Figure 7:** Average fit from the out-of-sample regressions each month sorted by year. Left panel, green squares: Adjusted  $R^2$ s from predictions made using an OLS model as in Equation (6). Left panel, blue triangles: Adjusted  $R^2$ s from predictions made using the LASSO as in Equation (10). Left panel, red circles: Adjusted  $R^2$ s from predictions made using both OLS and the LASSO as in Equation (12). 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 bad job of forecasting returns, and there are other years, such as 2010, where the OLS model does a 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 a factor of 1.5.”

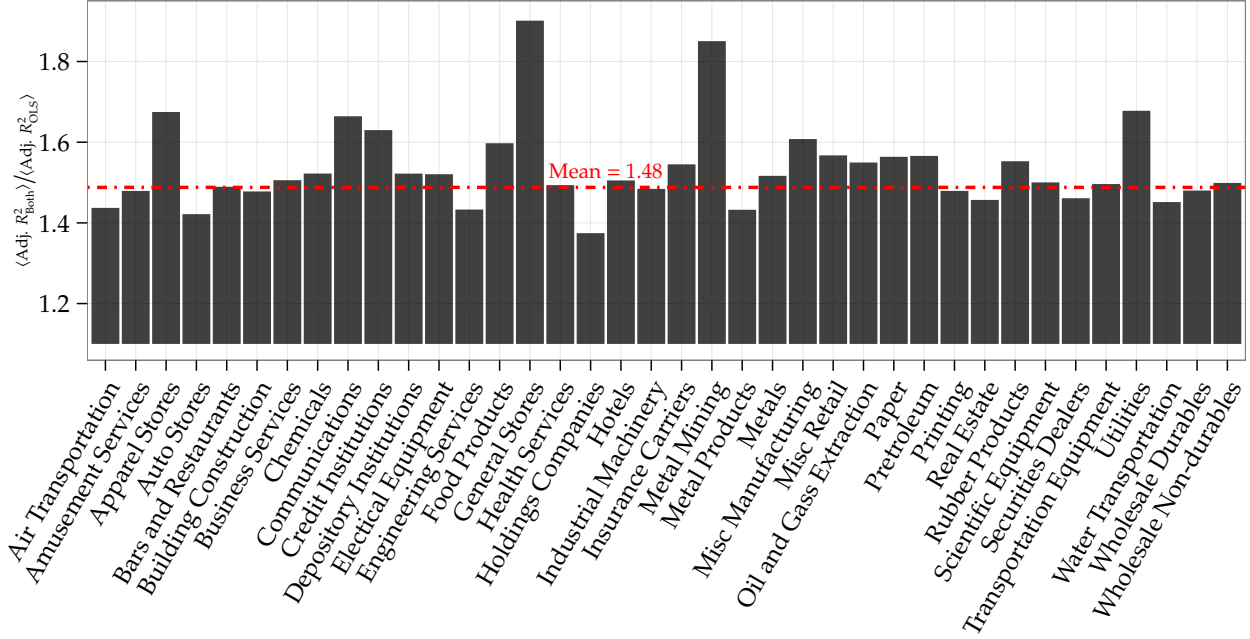
Percent?

the forecasting regressions each year. There are some years, such as 2007, where the OLS model does a bad job of forecasting returns (adjusted  $R^2 = 3.72$ ), and there are other years, such as 2010, where the OLS model does a good job of forecasting returns (adjusted  $R^2 = 9.32$ ). But, regardless of the average fit for the OLS model in any given year, the right panel of Figure 7 shows that including the LASSO’s return-forecast always boosts out-of-sample predictive power by a factor of 1.5.

*Industry Groupings.* Motivated by the evidence of industry lead-lag effects documented in both [Hong, Torous, and Valkanov \(2007\)](#) and [Hou \(2007\)](#), we show that the 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 8 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 gain factor of 1.5 is remarkably steady across all major industry groups.

*Market and Industry Returns.* We know that individual stock returns are explained by changes in market-wide and industry-specific returns. So, one potential concern is that our LASSO return-forecast is just picking up information that’s already in these variables. Put another way, perhaps it’s the case that the LASSO tends to make positive return forecasts for all stocks in a given minute, right before the whole market does really well. Or, maybe the LASSO tends to make positive return forecasts for all stocks in an industry at the same

## Distribution of Adjusted $R^2$ s, by Industry



**Figure 8:** 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.”

time, right before the industry does really well?

To show that this is not what’s going on, we include contemporaneous market-wide and industry-specific returns in our forecasting regressions,

$$r_{n,t+1} = \tilde{a}_n + \dots + \tilde{d}_{Mkt,n} \cdot r_{Mkt,t+1} + e_{n,t+1} \quad (13a)$$

$$\text{and } r_{n,t+1} = \tilde{a}_n + \dots + \tilde{d}_{Mkt,n} \cdot r_{Mkt,t+1} + \tilde{d}_{Ind(n),n} \cdot r_{Ind(n),t+1} + e_{n,t+1}, \quad (13b)$$

where “ $\dots$ ” denotes some combination of forecasting variables from Equation (6), (10), or (12) and the variables  $\tilde{d}_{Mkt,n}$  and  $\tilde{d}_{Ind(n),n}$  are the estimated coefficients for each stock on the contemporaneous market-wide return or its contemporaneous industry-specific return. Because we need to be able to estimate an industry-specific return, we restrict the sample to stocks in 3-digit SIC-code industries with at least 10 stocks in a given month.

Table 2 shows the summary statistics from these new regressions. The first, fourth and seventh columns correspond to the left-3 columns of Table 1 when restricting the sample to stocks in 3-digit SIC-code industries with at least 10 stocks in a given month. The coefficients are almost identical, suggesting that this restriction isn’t altering the sample in a meaningful way. From this starting point, we then add in controls for

the market-wide and industry-specific returns for each stock. As we do so, the point estimates in the regressions hardly change. What’s more, there is still a significant jump in the model’s fit from the third column (OLS return-forecast plus market-wide and industry-specific controls, average adjusted  $R^2 = 12.82\%$ ) to the ninth column (both the OLS and the LASSO return-forecasts plus market-wide and industry-specific controls, average adjusted  $R^2 = 15.36\%$ ). Thus, the LASSO’s return-forecast isn’t just a proxy for market-wide or industry-specific news. It’s capturing some other information.

*Penalty Parameters.* Finally, we find that our results are robust to selecting the penalty parameter,  $\lambda$ , in different ways. We select the  $\lambda$  for each stock by choosing the penalty parameter with the highest out-of-sample  $R^2$  during the first 45 minutes 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 isn’t necessarily optimal; but, we simply want to show that accounting for sparse signals can significantly boost traders’ out-of-sample predictive power. This procedure is the method of choice in [Friedman, Hastie, and Tibshirani \(2010\)](#). The right panel of [Figure 4](#) shows that the LASSO’s predictions do not depend on the gritty details of how  $\lambda$  is chosen.

### Out-of-Sample Return Predictability, Market and Industry Controls

	Full Sample								
$\langle \tilde{a}_n \rangle \times 10^{-4}$	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)	0.01 (18.90)
$\langle \tilde{b}_n \rangle \times 10^{-4}$	3.60 (135.83)	3.61 (136.46)	3.51 (137.91)				3.02 (131.45)	3.04 (132.12)	2.95 (133.35)
$\langle \tilde{c}_n \rangle \times 10^{-4}$				3.18 (160.75)	3.18 (161.28)	3.09 (164.21)	2.40 (170.42)	2.40 (170.95)	2.34 (174.42)
$r_{Mkt,t+1}$		✓	✓		✓	✓		✓	✓
$r_{Ind(n),t+1}$			✓			✓			✓
$\langle \text{Adj. } R^2 \rangle$	5.43%	10.88%	12.82%	4.54%	9.96%	11.96%	8.06%	13.51%	15.36%

**Table 2:** Average of the parameter estimates from the out-of-sample regressions each month described by Equations (6), (10), and (12) when including contemporaneous market-wide and industry-specific returns. The full sample includes results from (stock,minute)-level regressions in each October from 2005 to 2013 for stocks in 3-digit SIC-code industries with at least 10 stocks in a given month. Coefficient estimates have units of percent per minute. Numbers in parentheses are the  $t$ -statistics. Reads: “When we add controls for market-wide and industry-specific returns, the point estimates in the regressions hardly change and there is still a significant jump in the model’s fit from the third column (average adjusted  $R^2 = 12.82\%$ ) to the ninth column (average adjusted  $R^2 = 15.36\%$ ).”

### 3 Trading-Strategy Returns

Next, to show that this predictability isn't just a statistical artifact, we compute the returns to a trading strategy that buys or sells a stock whenever the LASSO's return-forecast exceeds the bid-offer spread. This plain-vanilla strategy generates returns of 0.30% per month net of trading costs, and these positive net returns exist in each subsample of the data we look at.

#### 3.1 Realized Returns

Table 1 shows 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 (see Moskowitz, Ooi, and Pedersen (2012)) is  $(390 \cdot 21) \cdot (3.17 \times 10^{-4}) = 2.60\%$  per month. 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 (10), we used information from future trading periods. For instance, the strategy dictated by Equation (10) 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

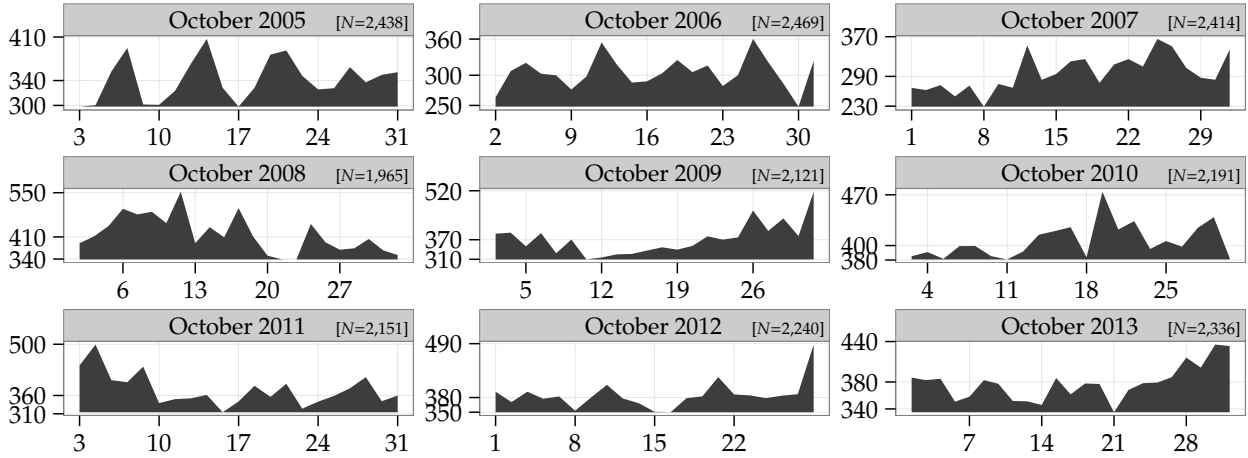
Trading-Strategy Returns

	Full Sample	2005-07	2008-10	2011-13
No Spread	2.82 (128.47)	2.28 (74.36)	2.97 (87.76)	3.28 (69.65)
NBBO Spread	0.30 (23.58)	0.17 (8.81)	0.17 (8.10)	0.55 (22.83)

**Table 3:** Returns per month to a trading-strategy based on LASSO return-forecast. In the first row, the strategy buys or sells a stock whenever the normalized LASSO's return-forecast is positive or negative, and no spread is paid. In the second row, the 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. The columns marked "2005-07", "2008-10", and "2011-13" report the same calculations done in 3 subsamples. Numbers in parentheses are *t*-statistics. Reads: "The trading strategy based on LASSO return-forecast generates a 0.30% per month return net of trading costs when applied to an average stock."



## Spread-Beating Returns per Minute



**Figure 9:** Daily average number of stocks with spread-beating returns,  $|r_{n,t+1}| > \text{spread}_{n,t}$  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.”

time-weighted

out-of-sample LASSO predictions for each stock,

$$\hat{\mu}_n^{\text{LASSO}} = \frac{1}{2,930} \cdot \sum_{t=1}^{2,930} f_{n,t}^{\text{LASSO}} \quad (14a)$$

$$\text{and } \hat{\sigma}_n^{\text{LASSO}} = \left( \frac{1}{2,930} \cdot \sum_{t=1}^{2,930} \left[ f_{n,t}^{\text{LASSO}} - \hat{\mu}_n^{\text{LASSO}} \right]^2 \right)^{1/2}. \quad (14b)$$

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

$$r_{n,t+1}^{\text{LASSO}} = \left( \frac{f_{n,t}^{\text{LASSO}} - 0}{\hat{\sigma}_n^{\text{LASSO}}} \right) \times r_{n,t+1}, \quad (15)$$

for the second half of each October from 2005 to 2013—that is, days 11 through 21. 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 9 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 utterly useless because it work for the  $364/2,260 \approx 16\%$  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:

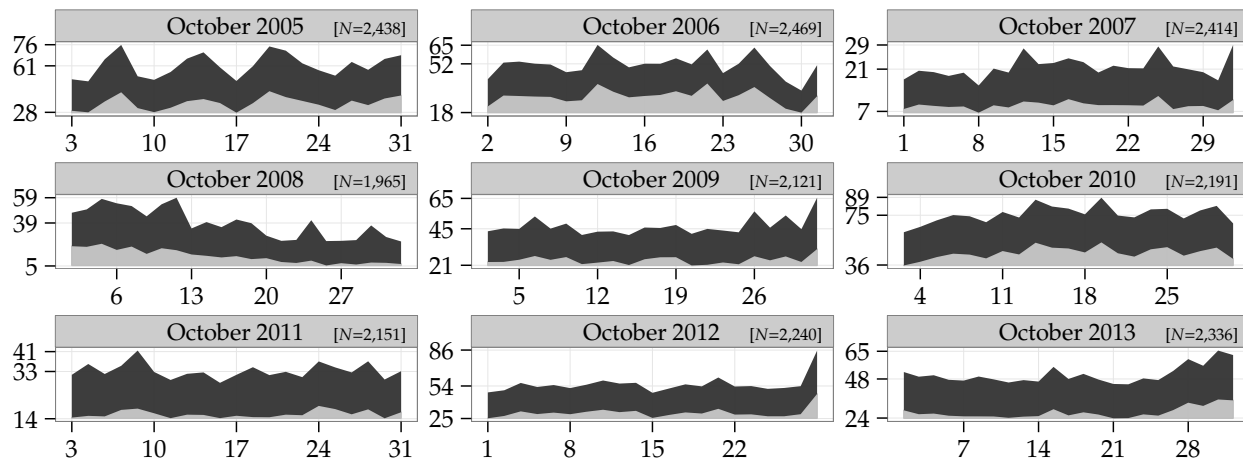
$$r_{n,t+1}^{\text{LASSO}} = \left\{ \left| \frac{f_{n,t}^{\text{LASSO}} - 0}{\hat{\sigma}_n^{\text{LASSO}}} \right| \times \left( \text{sgn}[f_{n,t}^{\text{LASSO}}] \cdot r_{n,t+1} - \text{spread}_{n,t} \right) \right\} \cdot 1_{\{|f_{n,t}^{\text{LASSO}}| > \text{spread}_{n,t}\}}. \quad (16)$$

So, for example, if  $\text{spread}_{n,t} = 0$ , then there is no spread and the strategy is the same as before. By contrast, if the spread is positive,  $\text{spread}_{n,t} > 0$ , then the trading strategy only invests when the LASSO's return-forecast is sufficiently large,  $|f_{n,t}^{\text{LASSO}}| > \text{spread}_{n,t}$ . 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,  $|r_{n,t+1}| > \text{spread}_{n,t}$ . For each stock, we define the spread as the average of the NBBO spread each minute in the second-by-second TAQ data.

*Estimation Results.* Table 3 describes the returns per minute to trading strategies based on the LASSO's return-forecasts under two different regimes: no spread and NBBO spread. The top row reveals that, for a typical NYSE-listed stock, the LASSO-based strategy generate positive gross returns of 2.82% per month in the absence of any trading costs. This point estimate is very close to the  $(390 \cdot 21) \cdot (3.17 \times 10^{-4}) = 2.60\%$  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 0.30% per month when applied to a typical stock. But, this return is still positive and statistically significant. What's more, the results exist in each subsample of the data we look at and are slightly increasing over time. 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 LASSO Predictions per Minute



**Figure 10:** 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.”

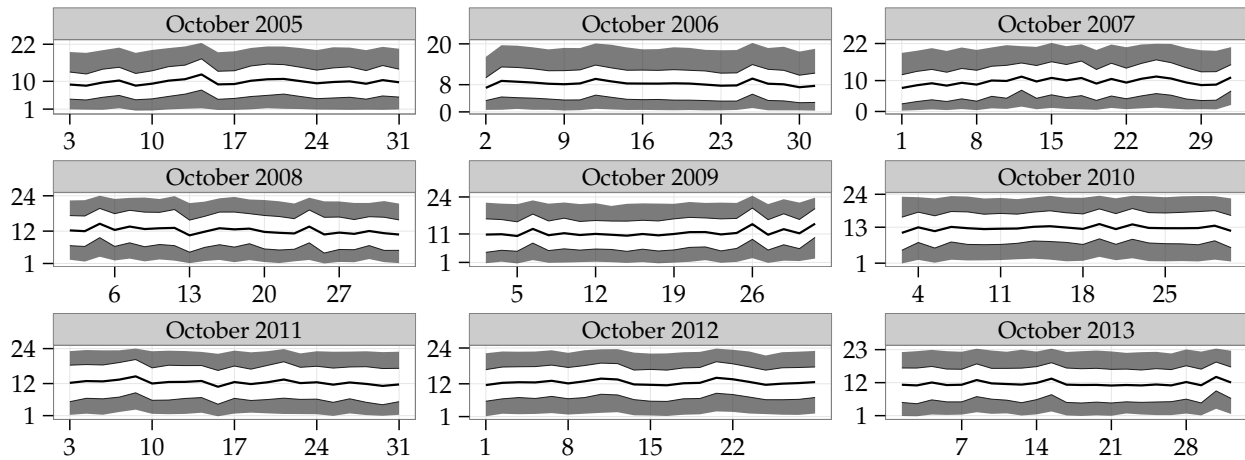
### 3.2 Different Trades

While the main goal of looking at the trading-strategy returns is to show that the LASSO’s predictive power is economically meaningful, this analysis also has the added benefit of giving us another vantage point for seeing how the information captured by the LASSO differs from the information captured by an OLS regression.

*Accurate Predictions.* Figure 10 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,  $|r_{n,t+1}| > spread_{n,t}$ , and where the LASSO said the stock would realize a spread-beating return,  $|f_{n,t}^{LASSO}| > spread_{n,t}$ . 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

## Significant Predictors per Stock per Minute



**Figure 11:** 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."

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.

## 4 Evidence of Sparsity

We've just seen that including the LASSO's return-forecast in a predictive regression boosts the out-of-sample  $R^2$  by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ , and that trading on the LASSO's return-forecast generates net returns of 0.30% per month. When we dig a little deeper to better understand where this predictive power comes from, we find that the LASSO's predictive power comes from identifying the right variables at the right time, not from better estimating the effect of some persistent factor.

### 4.1 Number of Predictors

To start with, the LASSO uses an extremely small number of predictors to make its return-forecast every minute.

*Predictors per Stock.* Figure 11 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

$$0.5\% = 11/2,191 \tag{17}$$

of the 2,191 possible stocks that the LASSO could choose from in October of 2010. Moreover, this is a stable feature of all the stocks we look at.

*Timeseries Variation.* The LASSO’s tendency to use only a handful 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 quite constant across our sample. The LASSO generally makes its return-forecast using only 11 predictors out of the roughly 2,000 possible each month.

## 4.2 Sparse Predictors

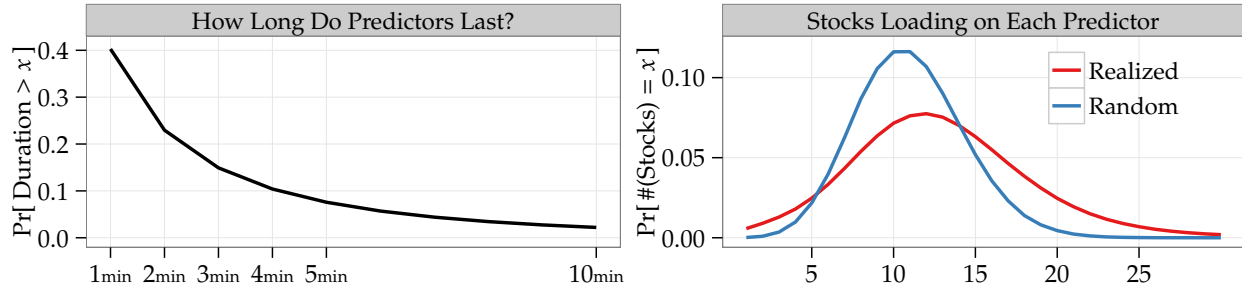
Even though the LASSO uses a small number of predictors when making its return-forecasts, this isn’t necessarily evidence that its identifying a sparse signal. It could be the case that the LASSO always chooses the same 11 predictors—in other words, that the signal is just a persistent factor. Or, it could be the case that the LASSO chooses entirely different predictors when forecasting each stock. We now show that neither of these two possibilities is true in the data.

*Predictor Duration.* To see what we mean, first notice that the LASSO’s chosen predictors do not remain significant for long. The left panel of Figure 12 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% 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 10 minutes later since the expected time until 11 failures is given by

$$7.33_{\text{min}} = \left( \frac{1-0.60}{0.60} \right) \times 11. \tag{18}$$

*Predictor Overlap.* Yet, in spite of the fact that the LASSO is constantly churning through predictors, we find that the predictors that the LASSO selects for each stock are far more

## Evidence of Sparse Predictors



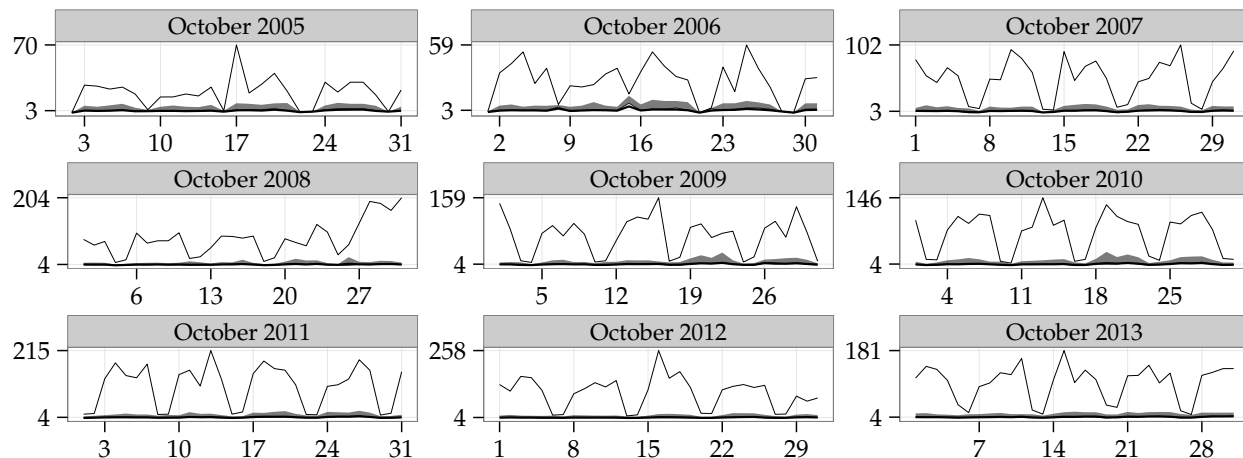
**Figure 12:** Left panel: Probability that the LASSO uses a predictor for more than  $x$  consecutive minutes when making the return-forecast for a particular stock. Reads: “Less than 10% of all LASSO predictors remain significant for more than 4 consecutive minutes.” Right panel, red line: Probability that the LASSO uses a particular predictor when making its return-forecast for  $x$  different stocks. Right panel, 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. Reads: “Compared to when each stock’s predictors are chosen at random, predictors are more likely to be used very often or not at all when the predictors are selected by the LASSO.”

likely to overlap than would be expected by pure chance. Put another way, if the LASSO is using the lagged returns of Family Dollar to forecast Chevron’s returns, then it is also much more likely to be using this variable when making return-forecasts for other stocks, as well. The red line in the right panel of Figure 12 gives the probability that a randomly selected stock is a significant predictor in  $x$  different LASSO return-forecasts. The blue line, on the other hand, gives the probability that a randomly selected stock would be a significant predictor for  $x$  return-forecasts if each stock had 11 randomly selected predictors. Compared to this random-selection benchmark, the LASSO is 17.6-times more likely to use a predictor in more than 20 of its return-forecasts than it would be by pure chance.

## 5 More Than Just News

We now data from RavenPack to show that the LASSO is doing more than just mirroring news announcements: it’s capturing how this news propagates from stock to stock. For example, even though Family Dollar is more likely to be chosen by the LASSO as a significant predictor for some other stock’s returns in the minutes following a news announcement about Family Dollar, this news announcement doesn’t reveal which stocks Family Dollar’s returns will help forecast. Should we look at oil and gas stocks like Exxon? Industrials like Mitsubishi? Somewhere else? The LASSO identifies these cross-stock links.

## News Releases Per Stock Per Day



**Figure 13:** 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.”

### 5.1 News-Release Data

We obtain business press data from RavenPack, a news analytics company.

*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, including [Barber and Odean \(2008\)](#), [Tetlock \(2010\)](#), and [Engelberg, Reed, and Ringgenberg \(2012\)](#). After collecting each news release, RavenPack also assigns it a relevance score that ranges from 0.0 (not relevant) to 1.0 (most relevant) and computes its subsequent news impact on the market on the same scale.

*Variable Definition.* We use the RavenPack data to compute a “has news” indicator variable for each stock in each minute of our sample,

$$hasNews_{n,t} = \begin{cases} 1 & \text{if there is a new release about stock } n \text{ in minute } t, \\ 0 & \text{else.} \end{cases} \quad (19)$$

We can also interact this variable with the relevant and impact variables provided by RavenPack. Because we only have relevance and impact data when there is a news release, we set these variables equal to zero in all other minutes.

## 5.2 Effect of News

We begin our analysis by verifying that the LASSO is more likely to use a stock as a predictor when it realizes a news release.

*Predictor Choice.* To set up our regressions, we define a pair of additional variables. First, we create another indicator variable at the stock-pair-by-minute level, which captures whether or not the  $n$ th stock was used by the LASSO to predict the  $m$ th stock's returns in minute  $(t + 1)$ ,

$$isUsed_{n \rightarrow m, t} = \begin{cases} 1 & \text{if LASSO uses stock } n \text{ to forecast stock } m\text{'s returns in } (t + 1), \\ 0 & \text{else.} \end{cases} \quad (20)$$

To make the definition concrete, recall that the LASSO makes its return-forecast for each stock using 11 predictors on average, so  $E\left(\sum_{m=1}^N isUsed_{n \rightarrow m, t}\right) = 11$ . But, we could also compute the sum the other way and ask how many times does the LASSO use the  $n$ th stock as a predictor in minute  $t$ ,

$$\#UsedBy_{n, t} = \sum_{m=1}^N isUsed_{n \rightarrow m, t}. \quad (21)$$

If the LASSO used the  $n$ th stock in all its return-forecasts, then  $\#UsedBy_{n, t} = N$ ; whereas, if the LASSO never used the  $n$ th stock in its return-forecast for any other stock, then  $\#UsedBy_{n, t} = 0$ .

*Regression Specifications.* To show that the LASSO is more likely to use a stock as a predictor when it has a news release, we run the regression below,

$$\#UsedBy_{n, t} = \tilde{a}_t + \tilde{b}_n + \tilde{c} \cdot hasNews_{n, t} + \epsilon_{n, t}, \quad (22)$$

where  $\tilde{a}_t$  and  $\tilde{b}_n$  are minute and stock fixed effects. If we were to estimate  $\tilde{c} = 1$ , for example, then this would mean that a stock with a news release in the current minute is typically used by the LASSO to make one additional return-forecast relative to its average. We include stock fixed effects because some firms realize more news releases than others. We include time fixed effects because there are some minute with many more news releases than others (see Figure 13).

In addition to this baseline specification, we also run specifications where the “has news” indicator variable is interacted with RavenPack’s measures of relevance and impact,

$$\#UsedBy_{n, t} = \tilde{a}_t + \tilde{b}_n + \tilde{c} \cdot \{hasNews_{n, t} \times newsRelevance_{n, t}\} + \epsilon_{n, t} \quad (23a)$$

$$\text{and } \#UsedBy_{n, t} = \tilde{a}_t + \tilde{b}_n + \tilde{c} \cdot \{hasNews_{n, t} \times newsImpact_{n, t}\} + \epsilon_{n, t}. \quad (23b)$$

The relevance variable is a measure of the extent to which the news release is about a



particular company. For example, a news release about the computer industry that briefly mentions Apple might have a relatively low relevance score; by contrast, a news release about the Apple iPhone would have a relevance score of 1.0. The news impact score is a forward looking measure, which measures the rise in market volatility in the 2 hours immediately after a news announcement. We run these additional regressions to verify that our specification is correct. We should expect that the estimated  $\tilde{c}$  should rise in both these regressions. More relevant news about a company should lead the LASSO to use it in more return-forecasts. And, if we know that the market gets really **volatility** right after a news release about a particular company, then the LASSO should be much more likely to use this company's stock as a predictor when forecasting other stocks' returns.

*Regression Results.* The first 3 columns of Table 4 show the results of these 3 regressions. In the first column, we see that the LASSO tends to use a stock in 0.65-more return-forecasts in the minute when it realizes a news announcement. This effect jumps to 0.88-more return-forecasts when we focus on news stories that are very relevant to the stock. Finally, we see in the third column that the LASSO is likely to use a stock in 2.01 additional return-forecasts if it had a news release that had a large effect on subsequent market volatility. These results give strong evidence that the LASSO is using stocks with news releases as predictors.

### 5.3 How Information Propagates

But, even though news announcements are good predictors of which stocks the LASSO will use as predictors, they say nothing about how the LASSO will use them.

*Regression Specification.* We encode this result in the following regression,

$$isUsed_{n \rightarrow m, t} = \tilde{a}_t + \tilde{b}_{n \rightarrow m} + \tilde{d} \cdot hasNews_{n, t} + \epsilon_{n \rightarrow m, t}, \quad (24)$$

which we run at the ordered-stock-pair-by-minute level. If we estimate  $\tilde{d} > 0$ , then this means that the LASSO is always more likely to use, say, Family Dollar to predict Exxon's returns when there is news about Family Dollar. We include time and ordered-stock-pair fixed effects for the same reasons as before.

*Regression Results.* The fourth column of Table 4 shows the results of this regression. We find a coefficient estimate of  $\tilde{d} = 0$ . This means that news announcements about a particular stock don't always propagate through the market in the same way. Sometimes news about Family Dollar will predict oil and gas companies' returns, but other times it will be relevant for industrials like Mitsubishi. Thus, the LASSO is doing more than just loading on news releases when they come out. It's telling you which other stocks this

news is relevant for. Put another way, it’s both identifying and estimating the relevant cross-stock signals.

## 6 Related Literature

The paper borrows from and brings together several strands of the statistics and empirical asset-pricing literatures.

*The LASSO.* To start with, this paper belongs to a growing literature applying the LASSO to econometric problems. For some examples, see [Belloni, Chen, Chernozhukov, and Hansen \(2012\)](#) and [Belloni, Chernozhukov, and Hansen \(2014\)](#). These papers answer the question of how to estimate treatment effects in an econometric setting where there are a large number of (potentially weak) instruments. [Hastie, Tibshirani, and Friedman \(2001\)](#) provide a general introduction to the LASSO and give the intuition behind the “[Betting on sparsity principle](#)”, which suggests you assume that the underlying truth is sparse and use an  $\ell_1$  penalty to try to recover it. If you’re right, you will do well. If you’re wrong—that is, if the underlying truth is not sparse—then no method can do well. [Meinshausen and Yu \(2009\)](#) gives an excellent overview of how well these LASSO-type estimators extend to settings with correlated right-hand-side variables. In addition to the LASSO, numerous other  $\ell_1$ -based penalized-regression techniques have been suggested in the statistics literature. For instance, consider the least-angle regression ([Efron, Hastie,](#)

More Than Just News

	#UsedBy <sub>n,t</sub>			isUsed <sub>n→m,t</sub>
<i>hasNews<sub>n,t</sub></i>	0.65 (8.84)			0.01 (0.21)
<i>hasNews<sub>n,t</sub> × newsRelevance<sub>n,t</sub></i>		0.88 (10.69)		
<i>hasNews<sub>n,t</sub> × newsImpact<sub>n,t</sub></i>			2.01 (13.20)	
$R^2$	93.2%	94.1%	94.5%	14.2%
Time FE	✓	✓	✓	✓
Group FE	✓	✓	✓	✓

**Table 4:** Coefficient estimates from the regressions in Equations (22), (23a), (23b), and (24). 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: “While the LASSO typically uses a stock with a news announcement in 0.65 additional predictions, you can’t predict which return-forecast the LASSO will use it for.”

Johnstone, and Tibshirani (2004)), the elastic net (Zou and Hastie (2005)), and the Dantzig selector (Candes and Tao (2007)).

*Economics of Sparsity.* Several other papers have also investigated the economic implications of sparsity. For instance, Gabaix (2014) introduces a sparsity-based model of bounded rationality where economic agents build simplified mental models of the world that are sparse. These agents use an  $\ell_1$ -type penalty to figure out which variables are of first-order importance. While sparse thinking is a useful heuristic for real-world situations, it is non-Bayesian unless the agent's decision problem exactly matches the statistical structure outlined in Park and Casella (2008). From the opposite perspective, Chincó (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.

*Return Predictability.* Finally, this paper relates to a long line of papers on momentum, return predictability, and information diffusion dating back to the early 1990s. These papers can be split into two categories: those that focus on same-stock predictability and those that focus on cross-stock predictability. Papers that use a stock's past returns to predict its future returns find negative autocorrelation at horizons shorter than 3 months (Fama (1965), Lo and MacKinlay (1990), and Conrad, Kaul, and Nimalendran (1991)), positive autocorrelation at horizons between 3 and 12 months (Jegadeesh (1990), Jegadeesh and Titman (1993), Asness (1994), Chan, Jegadeesh, and Lakonishok (1996), and Carhart (1997)), and negative autocorrelation at horizons longer than a year (De Bondt and Thaler (1985)). The cross-stock predictability literature finds lead-lag relationship in price movements across stocks (Lo and MacKinlay (1990) and Boudoukh, Richardson, and Whitelaw (1994)). In a closely related paper, DeMiguel, Nogales, and Uppal (2014) show that a vector-autoregression can capture some cross-stock signals.

There are numerous explanations for these return patterns. Hong and Stein (1999) and Hong, Lim, and Stein (2000) give and then test a theoretical model of slow-moving new to explain this pattern. Chordia and Subrahmanyam (2004) shows theoretically that market-maker inventory management is an important driver of short-term predictability, and Hendershott and Seacholes (2007) confirms this link empirically. Hasbrouck and Seppi (2001) investigates the common factors in prices, order flow, and liquidity. Harford and Kaul (2005) examines the importance of common factors in explaining order flow, returns, and trading costs, and they find that common factors are key drivers only for S&P 500 stocks. Coughenour and Saad (2004) show stocks handled by the same specialist firm show commonality in their liquidity. The one exception to this commonality literature is Tookes (2008), which derives a model to show that informed traders have incentives to

trade in the stock of competitors. Her model highlights the informational link between stocks which share the same product market.

## 7 Conclusion

This paper applies the least absolute shrinkage and selection operator (LASSO) to identify rare, short-lived, “sparse” signals in the cross-section of returns. We find that using the LASSO increases out-of-sample predictability in minute-by-minute NYSE returns by a factor of 1.5, from an adjusted  $R^2 = 5.43\%$  to an adjusted  $R^2 = 8.08\%$ , and generates trading-strategy returns of 0.30% per month net of trading costs. This predictive power comes from quickly identifying the right predictors at the right time, not from better estimating the effects of some persistent factor. The LASSO typically forecasts a stock’s returns using the lags of only 11 other stocks (a mere 0.5% of all possible choices), and 90% of these predictors last 4 minutes or less.

But, the LASSO’s return-forecast is helpful only under certain conditions as highlighted in the simulation-based analysis in Appendix A. When using rolling 30-minute windows to fit the model, the LASSO is only going to add predictive power if there are a few (that is, less than 30) important predictors with coefficients larger than  $\lambda$  in any 30-minute time window. But, if there are no significant predictors or if these signals are not sparse (that is, there are more than 30), then the LASSO’s return-forecast won’t be a helpful predictor. In the first case, there wouldn’t be any cross-stock signals to estimate. In the second case, there would be too many cross-stock signals to estimate using only 30 data points. It’s possible to bet on sparsity and lose.

So, the LASSO’s success suggests a new way of thinking about the economic forces behind stock returns. If the LASSO adds predictive power, then returns must contain a sparse component. If we only run OLS regressions like Equation (1), then it’s hard to think about anything other than persistent factors,  $\hat{\theta}_1 \cdot x_t$ , and idiosyncratic noise,  $\epsilon_{n,t+1}$ , driving stock returns. But, we know that factors like market and industry returns only account for roughly 20% of asset-return volatility (see [Campbell, Lettau, Malkiel, and Xu \(2001\)](#)). So, why do returns move around so much? Can the remaining 80% really just be due to the accidents of life? This paper suggests an alternative. The sparse shocks captured by the LASSO are more structured than noise but can’t be captured by OLS regressions.

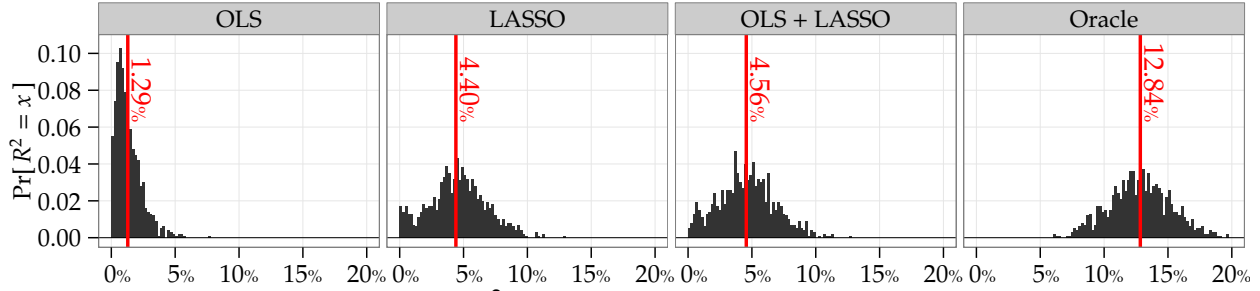
## References

- Asness, C. (1994). Variables that explain stock returns. Ph.D. Dissertation, University of Chicago.
- Barber, B. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies* 21(2), 785–818.
- 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. *The Review of Economic Studies* 81(2), 608–650.
- Boudoukh, J., M. P. Richardson, and R. Whitelaw (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *The Review of Financial Studies* 7(3), 539–573.
- Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *The Journal of Finance* 56(1), 1–43.
- Candes, E. and T. Tao (2007). The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *The Annals of Statistics* 35(6), 2313–2351.
- Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance* 52(1), 57–82.
- Chan, L., N. Jegadeesh, and J. Lakonishok (1996). Momentum strategies. *The Journal of Finance* 51(5), 1681–1713.
- Chinco, A. (2015). Feature-selection risk. SSRN Working Paper 2510536.
- Chordia, T. and A. Subrahmanyam (2004). Order imbalance and individual stock returns: Theory and evidence. *The Journal of Financial Economics* 72(3), 485–518.
- Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *The Journal of Finance* 63(4), 1977–2011.
- Conrad, J., G. Kaul, and M. Nimalendran (1991). Components of short-horizon individual security returns. *The Journal of Financial Economics* 29(2), 365–384.
- Coughenour, J. and M. Saad (2004). Common market makers and commonality in liquidity. *The Journal of Financial Economics* 73(1), 37–69.
- De Bondt, W. and R. Thaler (1985). Does the stock market over-react? *The Journal of Finance* 40(3), 793–805.

- DeMiguel, V., F. Nogales, and R. Uppal (2014). Stock return serial dependence and out-of-sample portfolio performance. *The Review of Financial Studies* 27(4), 1031–1073.
- Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *The Annals of Statistics* 32(2), 407–499.
- Engelberg, J., V. Reed, and M. Ringgenberg (2012). How are shorts informed?: Short sellers, news, and information processing. *The Journal of Financial Economics* 105(2), 260–278.
- Fama, E. (1965). The behavior of stock-market prices. *The Journal of Business* 38(1), 34–105.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *The Journal of Statistical Software* 33(1), 1–22.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics* 129(4), 1661–1710.
- Harford, J. and A. Kaul (2005). Correlated order flow: Pervasiveness, sources, and pricing effects. *The Journal of Financial and Quantitative Analysis* 40(1), 29–55.
- Hasbrouck, J. and D. Seppi (2001). Common factors in prices, order flows, and liquidity. *The Journal of Financial Economics* 59(3), 383–411.
- Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (1 ed.). Springer Science & Business Media.
- Hendershott, T. and M. Seacholes (2007). Market-maker inventories and stock prices. *The American Economic Review* 97(2), 210–214.
- Hong, H., T. Lim, and J. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance* 55(1), 265–295.
- Hong, H. and J. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance* 54(6), 2143–2184.
- Hong, H., W. Torous, and R. Valkanov (2007). Do industries lead stock markets? *The Journal of Financial Economics* 83(2), 367–396.
- Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *The Review of Financial Studies* 20(4), 1113–1138.
- Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The 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. *The Journal of Finance* 48(1), 65–91.
- Lo, A. and C. MacKinlay (1990). When are contrarian profits due to stock-market overreaction? *The Review of Financial Studies* 3(2), 175–205.

- Meinshausen, N. and B. Yu (2009). LASSO-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics* 37(1), 246–270.
- Moskowitz, T., Y. Ooi, and L. Pedersen (2012). Time-series momentum. *The Journal of Financial Economics* 104(2), 228–250.
- Park, T. and G. Casella (2008). The Bayesian LASSO. *The Journal of the American Statistical Society* 103(482), 681–686.
- Tetlock, P. (2010). Does public financial news resolve asymmetric information? *The Review of Financial Studies* 23(9), 3520–3557.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *The Journal of the Royal Statistical Society: Series B* 58(1), 267–288.
- Tookes, H. (2008). Information, trading, and product-market interactions: Cross-sectional implications of informed trading. *The Journal of Finance* 63(1), 379–413.
- Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *The Journal of the Royal Statistical Society: Series B* 67(2), 301–320.

## Adjusted $R^2$ Distribution: Simulated Data, Sparse Shocks



**Figure 14:** Distribution of adjusted  $R^2$ s from the forecasting regressions in Equations (6), (10), and (12) using simulated data generated from Equation (25). 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 as in Equation (6). Left-center panel: Out-of-sample prediction made using the LASSO as in Equation (10). Right-center panel: Out-of-sample predictions made using both OLS and the LASSO as in Equation (12). Far-right panel: Far-left panel: Out-of-sample prediction made using OLS as in Equation (6) 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.<sup>1</sup>

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

$$r_{q,t} = 0.15 \cdot \sum_{\mathcal{K}_t} r_{k,t-1} + 0.001 \cdot \epsilon_{q,t}, \quad (25)$$

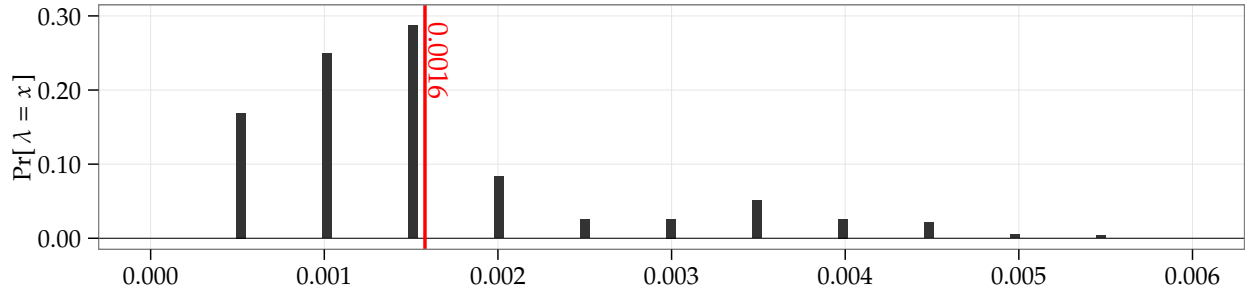
where  $\epsilon_{q,t} \stackrel{\text{iid}}{\sim} N(0,1)$ . This collection of  $K = 5$  sparse signals changes over time, leading to the time subscript on  $\mathcal{K}_t$ . We assume that there is a 1% chance that each signal changes every period, so each signal lasts  $(1 - 0.01)/0.01 = 99$  trading 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,  $q = 1$ , as defined in Equation (2) using the previous  $T = 50$  periods of data where the  $Q$  possible predictors are the  $Q = 100$  stocks. This means using  $T = 50$  time periods to estimate a model with  $Q = 100$  potential right-hand-side variables. As a useful benchmark, we also estimate the OLS model from Equation (1) and an oracle model. In this 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

<sup>1</sup>See <https://gist.github.com/alexchinco/467325abbf11d5c8f565>.



### λ Distribution: Simulated Data, Sparse Shocks



**Figure 15:** Distribution of the optimal choice of penalty parameters,  $\lambda$ , estimated using the first 150 trading periods in the simulated data generated from Equation (25). 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  $\lambda$ s 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.”

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 2.

*Forecasting Regressions.* We then check how closely these forecasts line up with the realized returns of asset  $q = 1$  by analyzing the adjusted  $R^2$  statistics from a bunch of forecasting regressions. For example, we take the LASSO’s return forecast in trading periods  $t = 151$  to  $t = 1,150$  and estimate the regression below,

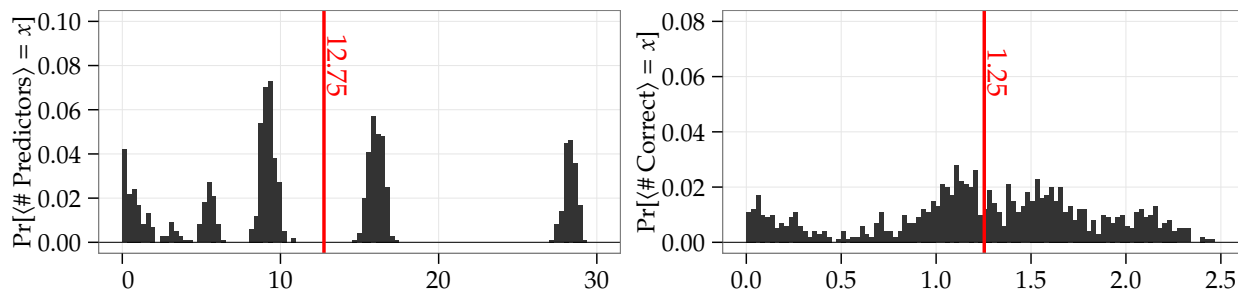
$$r_{1,t+1} = \tilde{a}_1 + \tilde{b}_1 \times \left( \frac{f_{1,t}^{\text{LASSO}} - \mu_1^{\text{LASSO}}}{\sigma_1^{\text{LASSO}}} \right) + \varepsilon_{1,t+1}, \quad (26)$$

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  $(t + 1)$ ,  $\mu_1^{\text{LASSO}}$  and  $\sigma_1^{\text{LASSO}}$  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 14 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,  $\lambda$ . 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 above only use data starting at  $t = 151$  instead of  $t = 51$ . Figure 15 shows the distribution of penalty parameter choices across the 1,000 simulations. The discrete 0.0005 jumps come from the discrete grid of possible  $\lambda$ s 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,

## Predictor Distribution: Simulated Data, Sparse Shocks



**Figure 16:** Distribution of the number of predictors used by the LASSO when making its return forecast using simulated data generated from Equation (25). 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.”

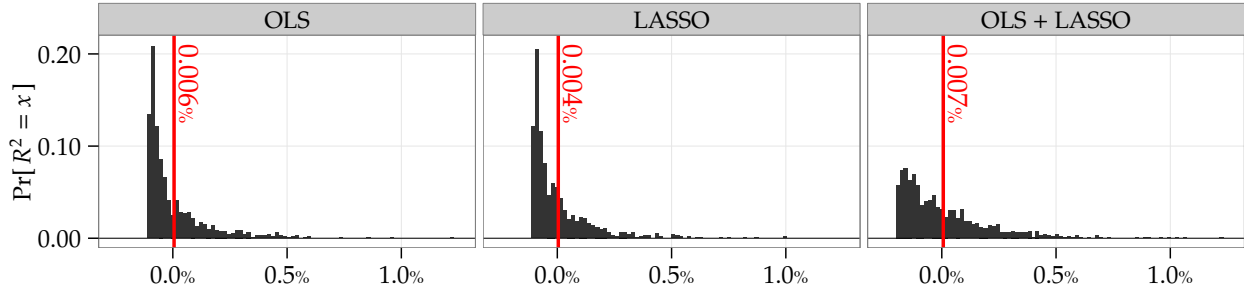
the left panel shows that the LASSO also locks onto lots of spurious signals. This result suggests that you might be able to improve the LASSO’s forecasting power by choosing a higher penalty parameter,  $\lambda$ .

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

$$r_{q,t} = 0.00 \cdot \sum_{k \in \mathcal{K}_t} r_{k,t-1} + \sigma \cdot \epsilon_{q,t}. \quad (27)$$

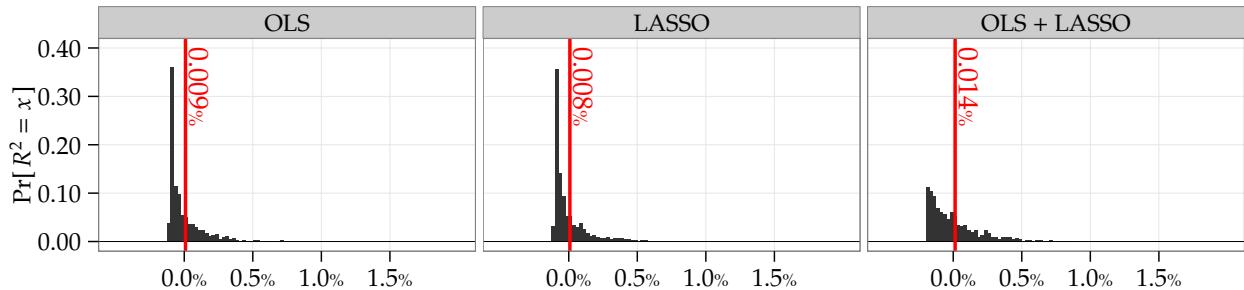
In the second setting, there are too many shocks:  $K = 75$ . Figure 17 and 18 show that, in both these settings, the LASSO doesn’t add any forecasting power. Thus, running these simulations offers a pair of nice placebo tests showing that the LASSO really is picking up sparse signals in the cross-section of returns.

### Adjusted $R^2$ Distribution: Simulated Data, No Shocks



**Figure 17:** Distribution of adjusted  $R^2$ s from the forecasting regressions in Equations (6), (10), and (12) using simulated data generated from Equation (27) 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 as in Equation (6). Center panel: Out-of-sample prediction made using the LASSO as in Equation (10). Right panel: Out-of-sample predictions made using both OLS and the LASSO as in Equation (12). Reads: “When there are no shocks, the LASSO does not add any forecasting power.”

### Adjusted $R^2$ Distribution: Simulated Data, Dense Shocks



**Figure 18:** Distribution of adjusted  $R^2$ s from the forecasting regressions in Equations (6), (10), and (12) using simulated data generated from Equation (25), 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 as in Equation (6). Center panel: Out-of-sample prediction made using the LASSO as in Equation (10). Right panel: Out-of-sample predictions made using both OLS and the LASSO as in Equation (12). Reads: “When shocks are dense, the LASSO does not add any forecasting power.”