SPARSE SIGNALS IN THE CROSS-SECTION OF RETURNS

ALEX CHINCO, ADAM D. CLARK-JOSEPH, AND MAO YE

ABSTRACT. Suppose that Hewlett-Packard’s stock price suddenly plunges. How quickly will IBM’s price reflect this news? This paper applies a statistical technique known as the least absolute shrinkage and selection operator (LASSO) to study how this kind of information flows from stock to stock. Using minute-by-minute NYSE returns from October 2010, we first show that accounting for such cross-stock information diffusion increases out-of-sample return predictability by 46.8%, from 6.95% to 10.20%. This means that a trading strategy which buys or sells a stock whenever the LASSO’s return-forecast for that minute exceeds the stock’s bid-ask spread generates a return of 0.16% per month net of trading costs. We then show that these cross-stock signals are sparse. Returns are typically predicted by the lags of only 13.7 other stocks (0.54% of all possible stocks), 90% of these predictors last for fewer than 5 minutes, and stocks tend to load on the same predictors at the same time.

JEL CLASSIFICATION. C55, C58, G12, G14

KEYWORDS. The LASSO, Sparsity, Return Predictability, Information Diffusion

Date: June 30, 2015.
University of Illinois at Urbana-Champaign. Chinco: alexchinco@gmail.com, (916) 709-9934. Clark-Joseph: adcj@illinois.edu, (217) 244-1536. Ye: maoye@illinois.edu, (217) 244-0474.

We have received many helpful comments and suggestions from John Campbell, Xavier Gabaix, Andrew Karolyi, Maureen O’Hara, Thomas Ruchti, Gideon Saar, Heather Tookes, and Sunil Wahal. 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.

Is it possible to predict a stock’s future returns using only past information? There are certainly a few select instances where the answer is “Yes.” For some examples, think about the medium-term momentum documented in Jegadeesh and Titman (1993), the long-run reversals in De Bondt and Thaler (1985), and the post-earnings-announcement drift in Ball and Brown (1968). However, these examples receive serious attention precisely because they are the exception rather than the rule. For the most part, the answer is “No.” Usually, when people check, prices tend to “reflect all available information” as Fama (1970) puts it.

But, let’s take a step back and think for a moment about how people check. The standard approach is to estimate an autoregression like the one below using ordinary least squares,

$$r_{IBM,t} = \hat{\theta}_0 + \hat{\theta}_1 \cdot r_{IBM,t-1} + \epsilon_{IBM,t},$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are estimated coefficients, $r_{IBM,t}$ denotes IBM’s return in minute $t$, and $\epsilon_{IBM,t}$ is the regression residual. The idea behind this test is that, if prices already reflect all publicly available information, then knowing IBM’s return in the current minute shouldn’t tell you much about what its return in the next minute will be—that is, $\hat{\theta}_1$ should be close to zero and the associated $R^2$ should be small. While this test is clearly a necessary condition for market efficiency, it completely ignores the relationships between stocks. For instance, suppose Hewlett-Packard’s stock price suddenly takes a plunge. Surely this signal contains some IBM-relevant information. Do we see this news immediately reflected in IBM’s share price? An autoregression simply cannot answer this question.

This paper proposes a technique that can: the least absolute shrinkage and selection operator (LASSO). When we apply the LASSO to NYSE returns, we find that accounting for cross-stock information diffusion increases out-of-sample return predictability by 46.8%.

Estimation Strategy. At this point, you might ask: why not just add lags of other NYSE-listed returns 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. There are 2,545 NYSE-listed stocks in our data for October 2010, so the resulting estimation problem would have 2,546 free parameters, one for the intercept and one for the return of each NYSE stock in the previous minute. Estimating all of these parameters in an autoregression using ordinary least squares (OLS) would then require a data sample that was at least 2,546 minutes—that is, 7 trading days. This is ludicrous. Predicting IBM’s return in the next minute using 7-day-old data isn’t just ill-advised, it’s archeology. To shorten the required sample length, we need to take a different approach and focus on only the most salient cross-stock links.

The LASSO allows us to do just that. This penalized-regression technique, which was introduced in Tibshirani (1996), estimates cross-stock predictability using far fewer time
periods by betting on sparsity—that is, by assuming only a handful of variables actually matter at any point in time. Using the LASSO to test whether IBM’s returns reflect all the information in the past returns of other NYSE stocks means solving the problem below,

$$\hat{\vartheta} = \arg \min_{\vartheta \in \mathbb{R}^{2,546}} \left\{ \frac{1}{2} \cdot T \cdot \sum_{t=1}^{T} \left( r_{IBM,t} - \vartheta_0 - \sum_{n=1}^{2,545} \vartheta_n \cdot r_{n,t-1} \right)^2 + \lambda \cdot \sum_{n=1}^{2,545} |\vartheta_n| \right\}, \quad (2)$$

where $r_{n,t-1}$ denotes the return on the $n$th stock in the previous minute, $\hat{\vartheta}$ is a $(2,546 \times 1)$-dimensional vector of estimated coefficients, $T$ denotes 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 in the curly brackets—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 it does this, consider the solution to Equation (2) when the right-hand-side variables are uncorrelated,

$$\hat{\vartheta}_n = \text{sgn}[\hat{\theta}_n] \cdot (|\hat{\theta}_n| - \lambda)_+,$$  

where $\hat{\vartheta}_n$ denotes the estimated coefficient from a standard OLS regression, $\text{sgn}[x] = x/|x|$, and $(x)_+ = \max\{0, x\}$. This solution means that, on one hand, if an unpenalized OLS regression would have estimated a large coefficient value, $|\hat{\theta}_n| \gg \lambda$, then the LASSO is going to deliver a similar estimate, $\hat{\vartheta}_n \approx \hat{\theta}_n$. On the other hand, if an OLS regression would have estimated a sufficiently small coefficient value, $|\hat{\theta}_n| < \lambda$, 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 even when the sample length is much shorter than the number of possible predictors, $T \ll 2,546$. Morally speaking, if only $K \ll 2,546$ 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

---

**Figure 1.** Average $R^2$ of out-of-sample return predictions using an autoregression, the LASSO, or both. Data: Minute-level returns for NYSE-listed stocks in October 2010. Reads: “Using the LASSO to predict returns boosts the out-of-sample $R^2$ by $(10.20 - 6.95)/6.95 = 46.8\%$. ”
these few important coefficients.

*Out-of-Sample Predictability.* When we apply the LASSO to minute-by-minute NYSE returns in October 2010, we find that these returns look substantially more predictable after accounting for cross-stock information diffusion. To provide a benchmark, we first estimate rolling autoregressions using 30 minutes of data and find that the average out-of-sample $R^2$ from these autoregressions is 6.95%. On average 6.95% 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. Next, we consider the effect of other stocks’ returns over the previous 3 minutes. This means using 30 minutes of data to both select and estimate a few significant predictors from among $1 + (3 \times 2,545) = 7,636$ possibilities. Our main result is that including the LASSO’s return-forecast boosts the out-of-sample $R^2$ by 46.8%, from 6.95% to 10.20%! Moreover, this predictive power is not driven by a few outlying observations—rather, it’s a robust feature of NYSE-listed stocks.

*Trading-Strategy Returns.* Next, we compute the returns to a trading strategy based on the LASSO’s return-forecast that buys or sells a stock whenever the LASSO’s return-forecast for that minute exceeds the stock’s bid-ask spread. This plain-vanilla strategy generates a return of 0.16% per month net of trading costs, and the strategy’s returns are largely independent of the accuracy of the benchmark autoregressive model. Put another way, the LASSO-based trading strategy generates positive net returns even when applied to stocks where the autoregressive model’s out-of-sample fit is extremely poor. These results provide supporting evidence that cross-stock information diffusion is important to real-world traders.

*Sparse Selection.* Although the trading-strategy returns are interesting in their own right, there is a deeper message hidden in the predictive power of the LASSO: many market signals are sparse. When we analyze how the LASSO is adding predictive power, we find that it selects only around 13.7 predictors on average. That is, the LASSO typically predicts a stock’s return in the current minute by considering the returns of only 13 or 14 other stocks at different points in time during the previous 3 minutes. To put these numbers in perspective, 13.7 stocks is only 0.5% of the 2,545 possible stocks that it can choose from. The LASSO is very picky. What’s more, stocks tend to load on the same predictors at the same time. We find that the LASSO is 3.3-times as likely to use a predictor in more than 20 of its return-forecasts than what would be expected by pure chance.

*Theoretical Implications.* This sparse structure has important implications for how we model traders’ decision problem. To put it bluntly, using autoregressions to test for market efficiency under-sells the difficulty of the problem they face. IBM’s current returns must not only reflect the information in its own past returns, but also all the available information about IBM hidden in past Hewlett-Packard-specific returns, past Adobe-specific returns, past Callaway-Golf-specific returns, and so on. Moreover, traders have to uncover these fleeting
Estimation Strategy and Timing

Step 1: Use a 30 minute sample period to estimate an AR(3) model.
\[ r_t = \hat{\theta}_0 + \hat{\theta}_1 \cdot r_{t-1} + \hat{\theta}_2 \cdot r_{t-2} + \hat{\theta}_3 \cdot r_{t-3} + \epsilon_t \]

Step 2: Use estimated coefficients and last 3 obs. to make out-of-sample prediction.

Figure 2. To make an out-of-sample predictions in minute \((t + 1) = 12:00pm\) using an autoregressive model, we estimate an autoregressive model 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 \((t + 1) = 12:00pm\), referring to the prediction as \( E_t[r_{t+1}] = f_{OLS}^{t} \). The figure above shows this process when the optimal number of lags is \( L^* = 3 \).

connections in real time. This is a much harder problem as highlighted in Chinco (2015).

2. Out-of-Sample Predictability

We apply the LASSO to minute-by-minute returns and find that they are 46.8% more predictable out-of-sample after accounting for cross-stock information diffusion. To do this, we use TAQ data on all NYSE stocks during October 2010, restricting the sample to the 2,545 stocks whose prices exceed $5 and which are actively traded every day of our sample.

2.1. Autoregression. To provide a benchmark, we begin by making out-of-sample predictions using an autoregression, an approach which explicitly does not take into account any cross-stock information diffusion. Figure 2 outlines the timing of the estimation strategy.

Coefficient Estimates. For each of the 2,545 NYSE-listed stocks in our TAQ data, we estimate separate autoregressions in rolling 30-minute windows,

\[ r_{n,t} = \hat{\theta}_0 + \sum_{\ell=1}^{L^*_n} \hat{\theta}_\ell \cdot r_{n,t-\ell} + \epsilon_{n,t}, \]  

where \( \hat{\theta}_0 \) and \( \{\hat{\theta}_\ell\} \) are estimated coefficients, \( L^*_n \) is the optimal number of lags for the stock during the 30-minute window according to the Akaike information criterion (AIC), \( r_{n,t} \) denotes stock \( n \)'s return in minute \( t \), and \( \epsilon_{n,t} \) is the regression residual. The first 30-minute window we consider each day is \( t \in \{10:16am, 10:17am, \ldots, 10:45am\} \) and the last window is
Choice of Tuning Parameters

![Graph](image)

**Figure 3.** Left panel: Fraction of the $3,780 \times 2,545 = 9,620,100$ different 30-minute time windows with 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 $R^2$ 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 $R^2$ for the prediction is 7.24%.”

$t \in \{1:15pm, 1:16pm, \ldots, 1:44pm\}$, yielding 180 total such samples for each stock on each day and $180 \times 21 = 3,780$ samples during the whole of October 2010. 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 3.

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

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

So, for example, if we estimated the coefficients for IBM over the 30-minute window $t \in \{11:30am, 11:31am, \ldots, 11:59am\}$, then we’d use these coefficients to predict IBM’s return in minute $(t + 1) = 12:00pm$. This gives us 180 out-of-sample predictions for each stock each day, one for each 30-minute training period, and a total of $180 \times 21 = 3,780$ different out-of-sample predictions for each of the 2,545 NYSE-listed stocks during October 2010.

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,

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

where $\tilde{\alpha}_n$ and $\tilde{\beta}_n$ are estimated coefficients, $r_{n,t+1}$ denotes stock $n$’s realized return in minute $(t + 1)$, $f_{n,t}^{OLS}$ denotes our prediction of stock $n$’s return in minute $(t + 1)$ using an autoregression model, $\mu_n^{OLS}$ and $\sigma_n^{OLS}$ 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. This means running a separate regression with 3,780 observations for each of the 2,545 NYSE-listed
Figure 4. Black bars: Probability that the $R^2$ from one of the 2,545 out-of-sample regressions falls within a 1%-point interval. Red vertical line: Average $R^2$ from these regressions. 
Left panel: Out-of-sample prediction made using an autoregression as in Equation (6). Middle panel: Out-of-sample prediction made using the LASSO as in Equation (11). Right panel: Out-of-sample predictions made using both an autoregression and the LASSO as in Equation (12). Reads: “While an autoregression generates out-of-sample predictions with an average fit of $R^2 = 6.95\%$, it has a fit of less than 1% for more than 17% of NYSE stocks.”

stocks in our sample.

The first column of Table 1 presents summary statistics describing the results of these 2,545 regressions. We find that, on average, stocks have a mean return each minute of zero to within one part in a million, $\langle \tilde{a}_n \rangle = 0.00 \times 10^{-4}\%$. Moreover, the gross monthly return to a market-timing strategy that is long a stock when the autoregressive model’s prediction is higher than average and short the stock otherwise is $(390 \cdot 21) \times \langle \tilde{b}_n \rangle = 3.41\%$ per month. In other words, $\tilde{b}_n$ is the average return per minute to a time-series momentum strategy à la Moskowitz, Ooi, and Pedersen (2012):

$$\tilde{b}_n = \frac{1}{T} \cdot \sigma_{n,\text{OLS}} \cdot \sum_{t=1}^{T} (\hat{r}_{n,t}^{\text{OLS}} - \mu_n^{\text{OLS}}) \cdot r_{n,t+1}. \tag{7}$$

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 also depends on knowing the distribution of the autoregressive 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 3.41% 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. As we should expect, this OLS predictor does not generate positive returns net of trading costs.

Although the regression coefficients are noteworthy, we are primarily focused on the regression fit. The first column of Table 1 shows that the average $R^2$ in these regressions is 6.95%, meaning that, for a randomly selected stock, you can explain 6.95% of the variation in its minute-by-minute returns using only information about that stock’s past returns. Table
LASSO vs. OLS Estimates

$$\hat{\theta}_{n,\ell}$$
$$\hat{\vartheta}_{n,\ell}$$
$$\lambda$$

Figure 5. x-axis: unpenalized regression coefficient in an infinite sample. y-axis: penalized regression coefficient. Dotted: OLS (i.e., unpenalized) regression coefficient corresponding to the $x = y$ line. Solid: LASSO coefficient. Reads: “If an OLS regression would have estimated a small coefficient value, $|\hat{\theta}_{n,\ell}| < \lambda$, then the LASSO sets $\hat{\vartheta}_{n,\ell} = 0$.”

1 also shows the results of this analysis for the first and second halves of our data separately, which indicate that these averages are stable across sample periods. If accounting for cross-stock information diffusion adds value, then this information has to allow us to improve on this 6.95% benchmark.

2.2. Penalized Regression. Let’s now consider the impact of other stocks’ returns over the previous 3 minutes. This means using 30 minutes of data to both select and estimate the few significant predictors from among $1 + (3 \times 2,545) = 7,636$ possibilities, a task that would clearly be impossible using a standard OLS regression. So, we use a penalized-regression approach known as the least absolute shrinkage and selection operator (LASSO).

Coefficient Estimates. For each of the 2,545 NYSE-listed stocks in our sample, we compute separate LASSO-based estimates in rolling 30-minute windows just like we did for the autoregressive approach. This means solving the optimization problem below,

$$\hat{\vartheta} = \arg \min_{\vartheta \in \mathbb{R}^{7,636}} \left\{ \frac{1}{2} \cdot 30 \cdot \sum_{t=1}^{30} \left( r_{n,t} - \vartheta_{n,0} - \sum_{n'=1}^{2,545} \sum_{\ell=1}^{3} \vartheta_{n',\ell} \cdot r'_{n',t-\ell} \right)^2 + \lambda \cdot \sum_{n'=1}^{2,545} \sum_{\ell=1}^{3} |\vartheta_{n',\ell}| \right\}, \quad (8)$$

where $\vartheta_{n,0}$ and $\{\vartheta_{n',\ell}\}$ are estimated coefficients, $r_{n,t-\ell}$ denotes stock $n$’s return $\ell$ minutes ago, and $\lambda$ is a penalty parameter. We perform this analysis on the exact same set of 3,780 different 30-minute training samples for each stock as in the autoregressions, starting each day with $t \in \{10:16am, 10:17am, \ldots, 10:45am\}$ and ending with $t \in \{1:15pm, 1:16pm, \ldots, 1:44pm\}$.

What additional information is the LASSO capturing? To answer this question, let’s take a look at the solution to the optimization problem in Equation (8) when the right-hand-side variables are uncorrelated,

$$\hat{\vartheta}_{n,\ell} = \text{sgn}[\hat{\theta}_{n,\ell}] \cdot (|\hat{\theta}_{n,\ell}| - \lambda)_+, \quad (9)$$

where $\hat{\theta}_{n,\ell}$ denotes the estimated coefficient from an OLS regression, $\text{sgn}[x] = x/|x|$, and $(x)_+ = \max\{0, x\}$. Equation (9) says that, if an OLS regression would have estimated a large coefficient value, $|\hat{\theta}_{n,\ell}| \gg \lambda$, then the LASSO is going to deliver a similar estimate,
\( \hat{\vartheta}_{n,t} \approx \hat{\vartheta}_{n,t} \). When you look all the way to the right or to the left in Figure 5, you see that the solid line denoting the LASSO estimate and the dotted line denoting the OLS estimate are quite close. By contrast, if an OLS regression would have estimated a sufficiently small coefficient value, \( |\hat{\theta}_n| < \lambda \), then the LASSO is going to pick \( \hat{\vartheta}_n = 0 \). This corresponds to the flat region in the solid line around the origin in Figure 5.

Thus, the LASSO’s return-forecast is helpful only under certain conditions. If there are only a few (that is, \( K < 30 \)) important cross-stock predictors with coefficients larger than \( \lambda \) in any 30-minute time window, then the LASSO will be able to estimate these links and provide useful information when trying to predict future returns. But, if there are no significant cross-stock predictors or if there is a large number of significant cross-stock predictors (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 links to predict. In the second case, there would be too many cross-stock links to characterize with only 30 data points. So, in other words, it’s possible to bet on sparsity and lose.

**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,

\[
E_t[r_{n,t+1}] = f^{\text{LASSO}}_{n,t} = \hat{\vartheta}_0 + \sum_{n=1}^{2,545} \sum_{\ell=1}^{3} \hat{\vartheta}_{n,\ell} \cdot r_{n,t-\ell+1},
\]

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

\[
r_{n,t+1} = \tilde{a}_n + \tilde{c}_n \cdot \left( \frac{f^{\text{LASSO}}_{n,t} - \mu_n^{\text{LASSO}}}{\sigma_n^{\text{LASSO}}} \right) + e_{n,t+1},
\]

where \( \tilde{a}_n \) and \( \tilde{c}_n \) are estimated coefficients, \( r_{n,t+1} \) denotes stock \( n \)’s realized return in minute \((t + 1)\), \( f^{\text{LASSO}}_{n,t} \) denotes our prediction of stock \( n \)’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. The second column of Table 1 displays summary statistics describing the results of these 2,545 regressions. We find that, for a typical stock, the average monthly return to a market-timing strategy that is long a security when the LASSO’s prediction is higher than average and short the security otherwise is \((390 \cdot 21) \times \langle \tilde{c}_n \rangle = 3.03\% \) per month, subject to the same implementation-related caveats as before. Looking at the average \( R^2 \) 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.

**Different Information.** Here is where things get interesting. When we include the out-of-sample predictions from both the autoregression and the LASSO in the same regression,

\[
r_{n,t+1} = \tilde{a}_n + \tilde{b}_n \cdot \left( \frac{f_{n,t}^{\text{OLS}} - \mu_n^{\text{OLS}}}{\sigma_n^{\text{OLS}}} \right) + \tilde{c}_n \cdot \left( \frac{f_{n,t}^{\text{LASSO}} - \mu_n^{\text{LASSO}}}{\sigma_n^{\text{LASSO}}} \right) + e_{n,t+1},
\]
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 2,545 combined regressions, reveals that including information from the LASSO increases the out-of-sample $R^2$ for the typical stock by $(10.29 - 6.95)/6.95 \approx 46.8\%$. What’s more, it turns out that for almost 20% of the stocks we analyze, the autoregressive benchmark does not provide a statistically significant contribution to forecasting, as shown by the large spikes around zero in the left-most panels of Figures 4 and 6. By contrast, LASSO offers statistically significant forecasting power for all but 2% of NYSE-listed stocks. Although the autoregressive model and the LASSO generate predictions with similar accuracies on average, each model uses very different information to make its forecast. The bet on sparsity pays off.

2.3. Robustness Checks. This pattern of return predictability is a robust feature of NYSE-listed stocks throughout our sample period. At the one-minute horizon, the autoregressive model and the LASSO tend to generate predictions with similar levels of accuracy using a variety of different ways.

Stable Over Time. Figure 7 shows how the predictive accuracy of the autoregressive model and the LASSO vary across each of the 21 trading days in October 2010. We find that, just like in the full sample, the autoregression’s prediction (denoted by the red bars) is more likely than the LASSO’s prediction (denoted by the blue bars) to be either extremely accurate or utterly worthless. On every single one of the trading days, the autoregression’s prediction is relatively more likely to have an $R^2 < 1\%$ or an $R^2 \geq 20\%$. Including the LASSO’s return-forecast when making out-of-sample predictions increases the prediction accuracy by at least

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>First Half</th>
<th>Second Half</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10/1 – 10/29</td>
<td>10/1 – 10/15</td>
<td>10/16 – 10/29</td>
</tr>
<tr>
<td>$\langle a_n \rangle$</td>
<td>$0.00 \times 10^{-4}$</td>
<td>$0.00 \times 10^{-4}$</td>
<td>$0.00 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(3.76)</td>
<td>(3.76)</td>
<td>(3.76)</td>
</tr>
<tr>
<td>$\langle b_n \rangle$</td>
<td>$4.16 \times 10^{-4}$</td>
<td>$3.54 \times 10^{-4}$</td>
<td>$4.00 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(44.23)</td>
<td>(41.77)</td>
<td>(46.61)</td>
</tr>
<tr>
<td>$\langle c_n \rangle$</td>
<td>$3.70 \times 10^{-4}$</td>
<td>$2.78 \times 10^{-4}$</td>
<td>$2.61 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(57.74)</td>
<td>(59.76)</td>
<td>(57.99)</td>
</tr>
<tr>
<td>$\langle R^2 \rangle$</td>
<td>$6.95%$</td>
<td>$10.20%$</td>
<td>$10.51%$</td>
</tr>
<tr>
<td></td>
<td>$5.64%$</td>
<td>$10.31%$</td>
<td>$7.25%$</td>
</tr>
<tr>
<td>Obs/n</td>
<td>3,780</td>
<td>1,980</td>
<td>1,800</td>
</tr>
</tbody>
</table>

Table 1. Average of the parameter estimates in the 2,545 different stock-specific regressions described by Equations (6), (11), and (12) when using the full sample covering all of October 2010 (left panel), the first half of the sample period (middle panel), and the second half of the sample period (right panel). Coefficient estimates have units of percent per minute. Numbers in parentheses are the t-statistics. Reads: “The typical fit of the LASSO’s prediction is on par with the typical fit of the autoregressive model’s prediction, 6.95% vs. 5.64%.”
Figure 6. Distribution of the slope coefficients in the 2,545 different stock-specific regressions described by Equation (12) when using the full sample covering all of October 2010. The vertical red lines represent the average of the slope coefficients and correspond to the point estimates \( \langle \tilde{b}_n \rangle \) and \( \langle \tilde{c}_n \rangle \) in the third column of Table 1. Reads: “While the autoregressive and LASSO models have similar coefficients on average, 3.54 \times 10^{-4} \text{ vs. } 2.78 \times 10^{-4}, they display very different cross-sectional distributions. For around 17\% of all the stocks the autoregressive model does not make a statistically significant out-of-sample prediction.”

Stable Across Industries. Motivated by the evidence of industry lead-lag effects documented in both Hong, Torous, and Valkanov (2007) and Hou (2007), we also show that the LASSO increases out-of-sample predictive power when we slice the data by industry. Table 2 displays the average \( R^2 \) values for the 2,545 different stock-specific regressions described by Equations (6), (11), and (12) for different industries, restricting the sample to the industries with at least 0.5\% of the stocks in our sample. The LASSO’s return-forecast adds significant predictive power to the out-of-sample regressions for all industry groupings.

Penalty Parameter Choice. Finally, we find that our results are robust to selecting the penalty parameter, \( \lambda \), in different ways. In our main specifications, 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 is then constant throughout the rest of the trading day. This is why our first prediction each day is 10:46am. Choosing the penalty parameter using the first 45 minutes of the trading day is in no way optimal; but, as we’ve discussed earlier, our goal is not to produce to the optimal forecast using the LASSO. We simply want to demonstrate that accounting for cross-stock information diffusion in a naïve way can significantly boost traders’ out-of-sample predictive power. This training-vs-test-sample procedure for selecting the penalty parameter is also the method of choice in Friedman, Hastie, and Tibshirani (2010). As a robustness check, we show in the right panel of Figure 3 shows that the general quality of the LASSO’s predictions do not depend on the gritty details of how \( \lambda \) is chosen.
3. Trading-Strategy Returns

Next, as an alternative way of quantifying the importance of cross-stock information diffusion, we compute the returns to a trading strategy based on the LASSO’s return-forecast.

3.1. Realized Returns. In Section 2 we pointed out that, in an idealized world, you could interpret the slope coefficients from the out-of-sample regressions in Equations (6) and (11) as the average returns to a time-series momentum strategy as studied in Moskowitz, Ooi, and Pedersen (2012). There are, however, two issues with this interpretation in practice: look-ahead bias and trading costs. We now analyze the returns to an analogous trading strategy that corrects for this pair of concerns.

Look-Ahead Bias. Using the return-forecasts from both the autoregressive model and the LASSO described in Sections 2.1 and 2.2 above, \( s \in \{\text{OLS, LASSO}\} \), we compute the returns to a trading strategy that is long whenever the prediction is positive and short whenever the
prediction is negative,

\[ r_{n,t}^s = \left( \frac{f_{n,t-1}^s - 0}{\sigma_n^s} \right) \times r_{n,t}, \quad (13) \]

for the second half of October 2010, \( t \in \{1981, 1982, \ldots, 3780\} \). We use the first 11 days of October 2010, \( t \in \{1, 2, \ldots, 1980\} \), to compute the volatility of the out-of-sample OLS and LASSO predictions for each stock:

\[ \sigma_n^s = \frac{1}{1980} \sum_{t=1}^{1980} f_{n,t}^s. \quad (14) \]

In other words, the trading strategy buys a stock whenever the out-of-sample prediction is positive, \( f_{n,t-1}^s > 0 \), and sells a stock whenever the out-of-sample prediction is negative.

## Out-of-Sample Regression \( R^2 \) by Industry

<table>
<thead>
<tr>
<th>Industry</th>
<th>( \langle R^2_{\text{OLS}} \rangle )</th>
<th>( \langle R^2_{\text{LASSO}} \rangle )</th>
<th>( \langle R^2_{\text{Both}} \rangle )</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banks</td>
<td>5.44%</td>
<td>4.81%</td>
<td>8.36%</td>
<td>93</td>
</tr>
<tr>
<td>Basic Materials</td>
<td>7.71%</td>
<td>6.04%</td>
<td>10.96%</td>
<td>137</td>
</tr>
<tr>
<td>Brokerage Svc.</td>
<td>2.04%</td>
<td>2.45%</td>
<td>3.90%</td>
<td>26</td>
</tr>
<tr>
<td>Consumer Goods</td>
<td>7.86%</td>
<td>6.33%</td>
<td>11.31%</td>
<td>181</td>
</tr>
<tr>
<td>Consumer Svc.</td>
<td>6.49%</td>
<td>6.00%</td>
<td>10.05%</td>
<td>187</td>
</tr>
<tr>
<td>Finance Companies</td>
<td>6.48%</td>
<td>4.95%</td>
<td>9.18%</td>
<td>28</td>
</tr>
<tr>
<td>Financial Svc.</td>
<td>6.96%</td>
<td>5.59%</td>
<td>10.14%</td>
<td>85</td>
</tr>
<tr>
<td>Gas Svc.</td>
<td>9.37%</td>
<td>6.97%</td>
<td>12.96%</td>
<td>21</td>
</tr>
<tr>
<td>Health Care</td>
<td>7.42%</td>
<td>6.94%</td>
<td>11.66%</td>
<td>100</td>
</tr>
<tr>
<td>Industrials</td>
<td>9.28%</td>
<td>7.53%</td>
<td>13.39%</td>
<td>308</td>
</tr>
<tr>
<td>Insurance</td>
<td>6.45%</td>
<td>6.20%</td>
<td>10.32%</td>
<td>100</td>
</tr>
<tr>
<td>Oil and Gas</td>
<td>7.00%</td>
<td>6.60%</td>
<td>10.97%</td>
<td>175</td>
</tr>
<tr>
<td>REITs</td>
<td>7.98%</td>
<td>7.21%</td>
<td>12.20%</td>
<td>134</td>
</tr>
<tr>
<td>Technology</td>
<td>7.37%</td>
<td>6.12%</td>
<td>10.74%</td>
<td>66</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>6.24%</td>
<td>6.14%</td>
<td>10.11%</td>
<td>59</td>
</tr>
<tr>
<td>Utilities</td>
<td>5.22%</td>
<td>6.66%</td>
<td>9.96%</td>
<td>110</td>
</tr>
<tr>
<td>Closed-End Funds, Specialty</td>
<td>7.87%</td>
<td>4.15%</td>
<td>9.86%</td>
<td>119</td>
</tr>
<tr>
<td>Closed-End Funds, Stock</td>
<td>10.60%</td>
<td>4.88%</td>
<td>12.60%</td>
<td>117</td>
</tr>
<tr>
<td>Closed-End Funds, Bond</td>
<td>4.88%</td>
<td>2.89%</td>
<td>6.57%</td>
<td>247</td>
</tr>
<tr>
<td>Trusts</td>
<td>1.61%</td>
<td>2.04%</td>
<td>3.16%</td>
<td>110</td>
</tr>
</tbody>
</table>

Table 2. Average \( R^2 \)'s in the 2,545 different stock-specific regressions described by Equations (6), (11), and (12) for different industries. The final column, “Obs.”, gives the number of NYSE-listed stocks in that industry. For brevity, we only report industries which make up at least 0.5% of the stocks in our sample. Reads: “The LASSO’s return-forecast adds significant predictive power to the out-of-sample regressions for all industry groupings.”
Trading-Strategy Returns

Panel (a): OLS Return-Forecast

<table>
<thead>
<tr>
<th></th>
<th>No Spread</th>
<th>Constant Spread</th>
<th>Stock-Spec. Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>5.05×10^{-4}</td>
<td>-0.30×10^{-4}</td>
<td>-0.65×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(18.18)</td>
<td>(1.38)</td>
<td>(8.39)</td>
</tr>
<tr>
<td>\langle R^2_{\text{OLS}} \rangle = H</td>
<td>8.18×10^{-4}</td>
<td>0.15×10^{-4}</td>
<td>-0.07×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(22.54)</td>
<td>(1.18)</td>
<td>(1.11)</td>
</tr>
<tr>
<td>4</td>
<td>7.65×10^{-4}</td>
<td>1.24×10^{-4}</td>
<td>-0.15×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(6.13)</td>
<td>(1.19)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>3</td>
<td>4.63×10^{-4}</td>
<td>-0.20×10^{-4}</td>
<td>-0.20×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(16.87)</td>
<td>(1.60)</td>
<td>(1.53)</td>
</tr>
<tr>
<td>2</td>
<td>3.25×10^{-4}</td>
<td>-0.80×10^{-4}</td>
<td>-0.98×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(11.55)</td>
<td>(6.96)</td>
<td>(6.49)</td>
</tr>
<tr>
<td>L</td>
<td>1.56×10^{-4}</td>
<td>-1.87×10^{-4}</td>
<td>-1.83×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(8.86)</td>
<td>(8.65)</td>
<td>(7.54)</td>
</tr>
</tbody>
</table>

Panel (b): LASSO Return-Forecast

<table>
<thead>
<tr>
<th></th>
<th>No Spread</th>
<th>Constant Spread</th>
<th>Stock-Spec. Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>4.36×10^{-4}</td>
<td>0.31×10^{-4}</td>
<td>0.19×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(28.28)</td>
<td>(3.38)</td>
<td>(3.67)</td>
</tr>
<tr>
<td>\langle R^2_{\text{LASSO}} \rangle = H</td>
<td>4.13×10^{-4}</td>
<td>0.28×10^{-4}</td>
<td>0.26×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(20.36)</td>
<td>(2.31)</td>
<td>(2.24)</td>
</tr>
<tr>
<td>4</td>
<td>4.27×10^{-4}</td>
<td>0.23×10^{-4}</td>
<td>0.20×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(23.97)</td>
<td>(3.09)</td>
<td>(2.77)</td>
</tr>
<tr>
<td>3</td>
<td>5.19×10^{-4}</td>
<td>0.41×10^{-4}</td>
<td>0.28×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(18.22)</td>
<td>(2.95)</td>
<td>(2.54)</td>
</tr>
<tr>
<td>2</td>
<td>5.54×10^{-4}</td>
<td>0.67×10^{-4}</td>
<td>0.16×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(9.30)</td>
<td>(1.73)</td>
<td>(1.90)</td>
</tr>
<tr>
<td>L</td>
<td>2.68×10^{-4}</td>
<td>-0.02×10^{-4}</td>
<td>0.05×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(9.68)</td>
<td>(0.14)</td>
<td>(0.28)</td>
</tr>
</tbody>
</table>

Table 3. Panel (a), All: Returns per minute to strategy based on OLS return-forecast when averaged across all 2,545 NYSE-listed stocks. Panel (a), \langle R^2_{\text{OLS}} \rangle = H, 4,3,2,L: Returns per minute to strategy based on OLS return-forecast when averaged across 5 different groups of stocks sorted by the out-of-sample fit of the OLS model during first 1,980 minutes. Panel (b), All: Returns per minute to strategy based on LASSO return-forecast when averaged across all 2,545 NYSE-listed stocks. Panel (a), \langle R^2_{\text{LASSO}} \rangle = H, 4,3,2,L: Returns per minute to strategy based on LASSO return-forecast when averaged across 5 different groups of stocks sorted by fit of the LASSO during the first 1,980 minutes. “Constant Spread”: Spread estimated as average spread for all 2,545 stocks during first 1,980 minutes. “Stock-specific Spread”: Spread estimated for each stock during first 1,980 minutes. Returns in percent per minute. Numbers in parentheses are t-statistics. 1,800 minute-specific observations per stock. Reads: “The trading strategy based on LASSO return-forecast generates a (390·21) × (0.19×10^{-4}) = 0.16% per month return net of trading costs when applied to an average stock.”
For a given prediction, the strategy dictates that we trade more in stocks where out-of-sample return-forecast is less volatile, \( \frac{\partial}{\partial \sigma_n} [f_{n,t-1} / \sigma_n^s] < 0 \). 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. By estimating each predictor’s volatility in an earlier period and setting each predictor’s mean to 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.

**Trading Costs.** Let’s now turn our attention to the second problem—namely, trading costs—which are substantial when trading every minute. Figure 8 highlights this basic point by showing that, out of the 2,545 NYSE-listed stocks in our sample, only around 114 realize returns in excess of their average estimated spread in any given minute. A predictor can be very good at forecasting small return fluctuations but still be utterly useless if it doesn’t also point out the \( \frac{114}{2,545} \approx 4.5\% \) 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 out-of-sample return-forecast exceeds 3-times the spread:

\[
 r_{n,t}^{s} = \left\{ \frac{f_{n,t-1}^s}{\sigma_n^s} - 0 \right\} \times \left( \text{sgn}[f_{n,t-1}^s] \cdot r_{n,t} - \text{spread}_n \right) \cdot \mathbb{1}_{\{|f_{n,t-1}^s| > 3 \cdot \text{spread}_n\}}. 
\]

So, for example, if \( \text{spread}_n = 0 \), then there is no spread and the strategy is the same as before. If the spread is positive, \( \text{spread}_n > 0 \), then the return-forecast has to have both the right sign as the realized return in the next minute, \( \text{sgn}[f_{n,t-1}^s] \cdot r_{n,t} > 0 \), and the realized return has to exceed the spread, \(|r_{n,t}| > \text{spread}_n\), for a trade to be profitable.

We consider two possible specifications for the spread, an average spread and a stock-specific spread, where both of these estimates are constant for each stock throughout the trading period. The stock-specific spread is computed as the average bid-ask spread for each
of the 2,545 stocks during the first half of October 2010—that is, a different spread for each stock. The average spread is average across the 2,545 stocks of the stock-specific spreads. In all cases, we demand that the return-forecast is at least 3-times as large as the spread before trading in order to avoid spurious predictions.

**Estimation Results.** Table 3 describes the returns per minute to trading strategies based on the OLS and LASSO return-forecasts using the portfolio weights given by Equation (15) under the three different trading-cost regimes. Moving from left to right across the top row of each panel, we see that, for a typical NYSE-listed stock in October 2010, both trading strategies generate positive gross returns, \((390 \cdot 21) \times 5.05 \times 10^{-4} \approx 4.14\%\) per month for the OLS-based strategy and \((390 \cdot 21) \times 4.36 \times 10^{-4} \approx 3.57\%\) per month for the LASSO-based strategy. As you would expect, introducing trading costs dramatically lowers both strategies’ returns, but they lower the returns to trading on the OLS’s return-forecast a lot more. After accounting for the spread, the OLS-based trading strategy has a negative net return of \((390 \cdot 21) \times -0.65 \times 10^{-4} \approx -0.53\%\) per month when applied to a typical stock; by contrast, the strategy based on the LASSO’s return-forecast still generates positive returns of \((390 \cdot 21) \times 0.19 \times 10^{-4} \approx 0.16\%\) per month. A more sophisticated LASSO-based strategy delivers much larger net returns.

The LASSO-based trading strategy’s returns are also largely independent of the accuracy of the benchmark autoregressive model. Table 4 shows that the LASSO-based trading strategy’s returns generates positive net returns even when applied to stocks where the autoregressive model’s out-of-sample fit is extremely poor. Its net returns beat those of the

<table>
<thead>
<tr>
<th>(\langle r_{n,t} \rangle)</th>
<th>(\langle R^2_{\text{OLS}} \rangle = L)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\langle R^2_{\text{OLS}} \rangle)</td>
<td>0.16x10^{-4}</td>
<td>0.14x10^{-4}</td>
<td>0.23x10^{-4}</td>
<td>0.19x10^{-4}</td>
<td>0.22x10^{-4}</td>
</tr>
<tr>
<td>(\langle r_{n,t} \rangle)</td>
<td>(1.20)</td>
<td>(1.86)</td>
<td>(1.43)</td>
<td>(2.12)</td>
<td>(2.41)</td>
</tr>
<tr>
<td>(\langle r_{n,t} \rangle - r_{n,t}^{\text{OLS}})</td>
<td>1.99x10^{-4}</td>
<td>1.12x10^{-4}</td>
<td>0.43x10^{-4}</td>
<td>0.35x10^{-4}</td>
<td>0.29x10^{-4}</td>
</tr>
<tr>
<td>(\langle r_{n,t} \rangle - r_{n,t}^{\text{OLS}})</td>
<td>(8.35)</td>
<td>(7.58)</td>
<td>(3.48)</td>
<td>(1.99)</td>
<td>(2.46)</td>
</tr>
</tbody>
</table>

**Table 4.** Average returns per minute (net of trading costs) during the second half of October 2010 for 5 different groups of stocks sorted by the fit of the autoregressive model during the first half of October 2010. \(\langle R^2_{\text{OLS}} \rangle = L, 4, 3, 2, \) and each of the 5 groups has \(2,545/5 = 509\) stocks. The spread is computed as the average spread for each stock during the first half of October 2010. \(\langle r_{n,t} \rangle\): Average returns per minute during the second half of October 2010 for the LASSO-based trading strategy. \(\langle r_{n,t}^{\text{OLS}} \rangle\): Average returns per minute during the second half of October 2010 for the OLS-based trading strategy. \(\langle r_{n,t} \rangle - r_{n,t}^{\text{OLS}}\): Average difference between the returns to a LASSO-based trading strategy and an OLS-based trading strategy during the second half of October 2010. Numbers in parentheses are t-statistics. 1,800 minute-specific observations per stock. Reads: “The point estimate for the average net returns per minute to the LASSO-based trading strategy is positive even for the quintile of stocks where the autoregressive model fits the worst—that is, in the column marked \(\langle R^2_{\text{OLS}} \rangle = L\).”
OLS-based trading strategy by \((390 \cdot 21) \times 1.99 \times 10^{-4} \approx 1.63\%\) per month in this lowest quintile. This is yet another way of showing that the LASSO is picking up different information than the autoregressive model and that cross-stock information diffusion is important to real-world traders.

3.2. Profitable Predictions. The LASSO captures important new information that the usual OLS-based techniques miss. This fact can be seen even more clearly by comparing and contrasting the spread-beating return predictions of each strategy.

Estimator Sensitivity. Figure 9 shows the number of stocks each minute that realize returns which exceed their spread, \(|r_{n,t+1}| > spread_n\), and which were accurately predicted by either the OLS return-forecast (top panel, \(|f_{n,t-1}^{OLS}| > 3 \cdot spread_n\)) or the LASSO return-forecast (bottom panel, \(|f_{n,t-1}^{LASSO}| > 3 \cdot spread_n\)). Light shaded: Average number of stocks that realized spread-beating returns and were accurately predicted by both the OLS and the LASSO each minute. Light shaded region is identical in both panels. Data: NYSE-listed stocks traded from 10:46am to 1:45pm during October 2010 averaged over each hour to make the plot more readable. Reads: “While the autoregressive model makes roughly 1 accurate prediction each minute during the sample, the LASSO only makes an accurate prediction once every 3 minutes. An accurate prediction is made by both models at the same time once every 11 minutes on average.”

Figure 9. Dark shaded: Average number of stocks each minute with spread-beating returns, \(|r_{n,t}| > spread_n\), that were accurately predicted by either the OLS return-forecast (top panel, \(|f_{n,t-1}^{OLS}| > 3 \cdot spread_n\)) or the LASSO return-forecast (bottom panel, \(|f_{n,t-1}^{LASSO}| > 3 \cdot spread_n\)). Light shaded: Average number of stocks that realized spread-beating returns and were accurately predicted by both the OLS and the LASSO each minute. Light shaded region is identical in both panels. Data: NYSE-listed stocks traded from 10:46am to 1:45pm during October 2010 averaged over each hour to make the plot more readable. Reads: “While the autoregressive model makes roughly 1 accurate prediction each minute during the sample, the LASSO only makes an accurate prediction once every 3 minutes. An accurate prediction is made by both models at the same time once every 11 minutes on average.”
Average Number of Inaccurate Predictions per Minute

3 · spread \( n \)) or the LASSO return-forecast (bottom panel, \(|f_{n,t-1}^{\text{LASSO}}| > 3 \cdot \text{spread}_n\)). These stocks represent trades where the strategies lost money. The first thing that jumps out from the figure is that the trading strategy based on the autoregressive model’s return-forecast makes nearly 6 times as many errors as the LASSO-based strategy. This is exactly what you’d expect since the LASSO explicitly penalizes weak predictors. A big reason why the OLS-based and LASSO-based trading strategies perform so differently is that they tend to make different kinds of errors.

4. Sparse Selection

We’ve just seen that returns are much more predictable after accounting for cross-stock information diffusion. Including the LASSO’s return-forecast in a predictive regression boosts the out-of-sample \( R^2 \) by 46.8% and trading on the LASSO’s return-forecast generates net returns of 0.16% per month. Although these findings are interesting in their own right, the LASSO’s success contains a deeper message: market signals are sparse—the bet on sparsity paid off. In any given minute, there are a few stocks that serve as important indicators for a large fraction of the market. For a trader trying to predict future returns, a big part of the challenge is just identifying the few relevant predictors in a timely manner. We now give some more direct evidence that market signals are sparse.

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. Looking at the red horizontal line, we see that on average the LASSO uses only 13.7 predictors to
make its return forecast. To put this number in perspective, note that this is roughly

$$0.5\% = \frac{13.7}{2,545}$$  \hspace{1cm} (16)

of the 2,545 possible stocks that the LASSO could possibly choose from. Moreover, this is a stable feature of all the stocks we look at. The histogram on the right-hand side of Figure 11 shows the distribution of the average number of predictors used for each stock and reveals that the LASSO uses between 9 and 18 predictors when making the return-forecast for more than 80% of all stocks.

**Timeseries Variation.** The LASSO’s tendency to use only a handful predictors is extremely stable over time, as shown in the left-hand side of Figure 11. 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 earlier in each trading day, the basic pattern is quite constant across our sample. The LASSO generally makes its return-forecast using only 13 or 14 predictors out of a possible 2,545.

4.2. **The Same Predictors.** In addition, even though we form our out-of-sample return-forecasts separately for each stock, the LASSO tends to pick out the same predictors for different stocks far more often than you’d expect to see by pure chance.

**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; and, less than 10% of all LASSO predictors remain significant for more than 4 minutes. If we estimate
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.”

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 used 14 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 13.7 failures is given by

$$9.33\text{min} = \frac{(1 - 0.60)}{0.60} \times 13.7.$$

**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 likely to overlap than would be expected by pure chance. 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 14 randomly selected predictors. Compared to this random-selection benchmark, the LASSO is 3.3-times more likely to use a predictor in more than 20 of its return-forecasts and 2.1-times more likely to use a predictor in fewer than 9 return-forecasts.

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

**Sparse Thinking.** Sparsity also has important economic implications. For instance, Gabaix (2014) introduces a sparsity-based model of bounded rationality where economic agents build simplified 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, 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, regardless of how smart they are.

**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 Whitney (1994)).

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.

6. Conclusion

Asset prices adjust to incorporate new information, and these price movements transmit information. While most of modern finance builds upon this basic notion, empirical studies of how this cross-stock information diffusion takes place are relatively scarce. Moreover, the few studies that we actually have focus primarily on mechanisms that operate outside the market, such as word-of-mouth communication (Hong, Kubik, and Stein (2005)), peer effects (Cohen, Frazzini, and Malloy (2008) and Cohen, Frazzini, and Malloy (2010)), or news reactions (Tetlock (2007)). But, how do real-world traders extract the relevant signals for forecasting a particular stock’s return from past prices, from the churning sea of recent market activity?

In principle, looking at the behavior of all other stocks’ past prices should suffice to disentangle the different components of complicated information, but this brute-force approach requires a large number of time-series observations using the standard econometric toolkit based on least-squares estimation. The world may move on before you can obtain enough observations to digest what the market is telling you. Since the market does somehow manage to process complicated information in a timely manner, this task must be tractable.

We suggest an alternative to the brute-force ordinary-least-squares approach: the least absolute shrinkage and selection operator (LASSO). This estimator allows traders to focus on only a small subset of stocks and dramatically reduces the length of the sample period. Using minute-by-minute NYSE returns from October 2010, we find that accounting for such cross-stock information diffusion increases out-of-sample return predictability by $46.8\%$. We then show that these cross-stock signals are sparse. Returns are typically predicted by the lags of only $13.7$ other stocks in any given minute—roughly $0.5\%$ of all possible predictors—and stocks tend to load on the exact same predictors at the exact same time. Our analysis hints at the full depth of the maxim that prices transmit information.
Predictors That Are Significant For At Least 2% of All Stocks

Figure 13. 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.
References


