Asset Insurance Premium in the Cross-Section of Asset Synchronicity

Raymond C. W. Leung and Yu-Man Tam*

Abstract

Any asset can use some portfolio of similar assets to insure against its own factor risks, even if the identities of the factors are unknown. A long position of an asset and a short position of this portfolio forms an *asset insurance premium* (AIP) that is different from the equity risk premium. We estimate the AIP by projecting a stock's return onto the entire asset returns span using a machine learning method. Stocks least (most) synchronized with other stocks earn a monthly AIP of 0.976% (0.305%). Asset synchronicity is countercyclical: high consumption growth correlates with low average asset insurance premium.

*Leung is with the Cheung Kong Graduate School of Business, and Tam is with the Office of the Comptroller of the Currency, U.S. Department of Treasury. Please send all correspondences to Leung at `raymondleung@ckgsb.edu.cn`. Part of this research is done while Leung was visiting UC Berkeley. We thank Robert M. Anderson, Alex Michaelides, Neal Stoughton, Baolian Wang, Hongjun Yan, and Wenhao Yang for helpful discussions. We also thank the seminar participants of the Five-Star Forum, Internal Machine Learning Workshop at the OCC, and SWUFE-CDAR 2018 Symposium for helpful comments and suggestions. All errors are ours.

Given any stock, how can one hedge against its factor risks? This question is simple to answer with a linear factor model structure. For instance, under the Markowitz mean-variance portfolio theory and its resulting equilibrium capital asset pricing model (CAPM) (Sharpe (1964), Lintner (1965)), any given stock's returns can be explained by some linear combination of a risk free asset return and its beta loading on the market portfolio return. Hence, the factor risk of any stock can be hedged out by shorting its beta loading multiplied by the market factor. However, with the large number of tradable and non-tradable factors that have been documented in the literature (Harvey, Liu, and Zhu (2016)), our question becomes difficult to approach because its answer then significantly depends on which factors the researcher decides to include in his empirical study, and is also affected by the empirical uncertainty with estimating the factor loadings.

In this paper, we first *theoretically* argue and then *empirically* show that an effective way to hedge against potentially unidentifiable factor risks of a stock is to answer this dual question: *Given any stock, what portfolio of all other stocks is most similar to it?* Suppose all stocks are exposed to the same set of linear factors but with heterogeneous factor loadings. If one can identify a portfolio of other stocks that is the most "similar", or *synchronized*, to it, then this portfolio is also exposed to similar factor loadings of that given stock. A short position on this portfolio of similar stocks will hedge against some of the factor risks of that given stock. Thus a long position on the stock, and a short position on this portfolio will expose the holder to any remaining factor risks of the given stock that cannot be completely hedged out by the other similar stocks. Moreover, this long-short position does not require the econometrician to know the true underlying factor structure of the economy. We *theoretically* prove, with minimal assumptions, that this long-short position exactly equates to residual factor risks of a given stock. Due to its hedging nature, we call the expected value of the returns of this long-short position an *asset insurance premium* for the given stock.

There remain *empirical* questions, despite the theoretical existence of this *asset insurance premium*. How does one empirically construct this similarity portfolio from thousands of all other stocks for a given stock? Is this long-short return positive-valued to deserve the name asset insurance "premium" for any given *average* stock? Is there a *cross-sectional* effect, whereby stocks that are "less synchronized" behave differently than stocks that are "more synchronized" with all other stocks? Are there any *macroeconomic* reasons to explain the existence of an average and cross-sectional asset insurance premium, if they empirically exist at all?

The empirical contributions in this paper are: (a) For each given stock, we use the *elastic-net* estimator (a machine learning method), to estimate and construct a sparse portfolio of "similar" stocks out of thousands of possible stocks in the market, and that the resulting regression R-squared can be viewed as the *asset synchronicity* of a stock; (b) We show that the unconditional value-weighted portfolio of stocks have a positive-valued asset insurance premium of 0.575% per month; (c) We show that in the cross-section, value-weighted portfolio of stocks that are least (most) synchronized with all other stocks in the market have an asset insurance premium of 0.976% (0.305%) per month, and the difference is highly statistically significant; and (d) We show that asset synchronicity is highly countercyclical (i.e. low during economic expansions, and high during economic recessions or market crises), where a 1% change in monthly consumption is associated with a -1.718% change in the cross-sectional asset insurance premium. We emphasize that we are *not* proposing a new factor, and our results complement the existing cross-sectional asset pricing literature. We are documenting a cross-sectional behavior — the asset insurance premium — that hitherto has not been explored in the empirical asset pricing literature. We provide evidence that the asset insurance premium is distinctly different from the well-known equity risk premium.

We need to apply a machine learning paper to empirically test our theoretical prediction out of technical necessity. Mechanically speaking, to find other similar stocks we simply want to "regress" a given stock's time series of returns onto the time series of returns of all other stocks. While this mechanical operation may sound simple, it cannot be tackled with traditional econometric methods because we hit upon a core hurdle of financial assets data. First and foremost, the traditional workhorse *ordinary least squares (OLS)* estimator is outright infeasible to answer this question. The *rank condition* in the OLS necessarily requires the regressor matrix to have more rows than columns. In our case, the regressor matrix would be concatenated columns of the time series of returns of all other stocks. At any given month, there are about $N+1 \approx 4,000$ tradable stocks in the US equity market. Yet, we only have about 60 years of post-World War II data. That means at frequencies that are of interest to empirical asset pricing researchers (e.g. monthly, quarterly, yearly), there simply are far too many stocks than there are time periods, meaning $T \ll N+1$.[1] In contrast, the *elastic-net* machine learning estimator proposed by Zou and Hastie (2005) is appropriate for our economic problem. The elastic-net is a generalization of the popular *LASSO* estimator of Tibshirani (1996). And indeed, the LASSO is itself a generalization of the OLS estimator. The elastic-net is well suited to answering our question for

[1]Cochrane (2005) summarizes the data nature of financial assets as: "Furthermore, even if one researcher is pure enough to follow the methodology of classical statistics, and wait 50 years for another fresh sample to be available before contemplating another model, his competitors and journal editors are unlikely to be so patient."

several reasons: (i) a regression problem with $T \ll N+1$ is perfectly feasible; (ii) while multicollinearity of asset returns is a serious issue for OLS, handling multicollinearity is a feature of the elastic-net due to its "grouping property"; and (iii) although the set of stocks $N + 1 \approx 4000$ is large, the "sparsity property" of the elastic-net implies only a small handful of stocks will explain the returns of a given stock.

We describe our estimation and "similar" portfolio construction method for each stock. At the end of each month, we collect the past twelve month's worth of daily return data of all stocks in the CRSP dataset. We use the elastic-net estimator to regress each stock i's daily returns onto the daily returns of the N other stocks. We collect the resulting estimated high-dimensional regressor coefficient vector and the R-squared scalar of this stock i. We repeat this procedure for all stocks $i = 1, \ldots, N+1 \approx 4000$. After the estimation procedure is complete, we use the large N-dimensional estimated coefficient vector of stock i as investment weights for the risky component of this portfolio for stock i, while the remaining other investment weights will go into the risk free asset. We call this portfolio the *ghost (portfolio) of stock* i, where the term "ghost" is to reminisce of a shadow projected by the given stock i's *actual* returns onto the returns span of all other stocks. Due to the sparsity property of the elastic-net estimator, it is actually the rule and not the exception that the estimated coefficient vector will have many zeros, which means there are only a handful number of risky stocks that enter into the ghost portfolio of stock i. By this construction method, the ghost portfolio is precisely then the sparse projection of the actual stock i's returns onto the span of all other stocks.

We use standard portfolio sort methods of the empirical asset pricing literature to estimate expected returns. We sort each stock i's elastic-net R-squared's into decile bins. Stocks with the lowest (highest) R-squared's have the lowest (highest) synchronicity with all other stocks. We consider both equal- and value-weighted portfolios within each bin. We hold the portfolios for one month, evaluate its returns performance, and then rebalance. Unlike conventional papers in the asset pricing literature, we now actually have *three* types of one-month ahead returns for a given stock: (i) the return of the *actual* stock i; (ii) the return of the *ghost* of stock i; and (iii) the *long-short* return between actual stock i and its ghost. The mean long-short return is the empirical estimate of the *asset insurance premium.* We study the monthly time series of these three types of portfolios from December 1975 to December 2017.

Our paper belongs to a growing literature of applying machine learning methods to study empirical asset pricing questions. Recent papers have applied variants of the *least absolute shrinkage and selection operator (LASSO)* estimator of Tibshirani (1996). Feng, Stefano, and Xiu (2017) takes advantage of the sparsity property of LASSO and develops a multi-step approach to evaluate the price of risk of a given new factor above and beyond an existing set of factors. Freyberger, Neuhierl, and Weber (2017) uses adaptive group LASSO to select characteristics that provides marginal information for the cross section of expected stock returns. The literature has documented a large set of factors or characteristics (Harvey, Liu, and Zhu (2016)), and it is hoped that machine learning methods can substantially shrink down the number of factors that can explain the cross-section of returns. Chinco, Clark-Joseph, and Ye (2018) uses the LASSO to predict one-minute ahead return using lagged high frequency returns of other stocks as regressors. Gu, Kelly, and Xiu (2018) applies an extensive battery of machine learning methods and discover that such methods improves predictability accuracy over traditional methods. In contrast to the existing papers that apply machine learning methods in empirical finance, we directly use machine learning methods to construct new characteristics (e.g. *asset synchronicity* and *asset insurance premium*) that are motivated by an economic theoretical prediction, and show they have hitherto undocumented asset pricing implications. The literature has used regression R-squared's as a measure of asset synchronicity, but mostly in an international finance context. Morck, Yeung, and Yu (2000) is one of the first paper to show that stock prices in poor economies comove more than rich economies, and the authors attribute the difference the level of property rights protection. Jin and Myers (2006) relates stock price synchronicity to corporate governance issues in an international context.

Our paper is also related to the literature of pairs trading and substitutability of risky assets. [2] Gatev, Goetzmann, and Rouwenhorst (2006) finds pairs of similar stocks using the minimal distance between normalized historical prices, and argues that the resulting pairs trading strategy generates abnormal returns and also argues that the source of this profit is the mispricing of close substitutes. Krauss (2017) is a recent survey of the pairs trading literature. Wurgler and Zhuravskaya (2002) similarly also argues that stocks without close substitutes are likely to have large mispricings. By using a machine learning method in this paper, a stock need not be paired with just one additional stock; indeed, the closest substitute of a given stock could potentially be hundreds of all other stocks.

Section 1 lays out the theoretical framework of the paper. Section 2 explains our estimation methodology. The main results of the paper are in Section 3, where we empirically show an unconditional average asset insurance premium exists, and that the asset insurance premium also exists in the cross-section of

[2] We thank an anonymous conference reviewer for pointing out this connection to us.

asset synchronicity. Section 4 shows that asset synchronicity is highly countercyclical to macroeconomic fundamentals. Section 5 compares the elastic-net approach to a simpler Fama-French motivated portfolio construction method, and we fail to see any cross-sectional results with the latter method. Section 6 shows that the asset insurance premium is distinct from the well-known equity risk premium. Section 7 shows that the asset insurance premium exists in the cross-section of asset synchronicity even controlling for various known priced factors. We conclude in Section 8. All proofs are in Section A (Appendix) and there are some additional results and discussions in Section B (Internet Appendix).

1 Asset insurance premium: Theoretical motivation

We first prove a general theoretical asset pricing result that will guide our empirical research design.

Theorem 1.1. *Suppose there are $N+1$ risky assets, a single risk free asset, and all of these risky assets are governed by a linear factor structure with K number of factors with returns $\boldsymbol{F}$ as in* (1a)*. Suppose there are strictly more risky assets than factors, so $N > K$.*

Then the excess returns R_i of any individual risky asset i can be expressed as a linear combination of other risky asset returns as in (1b)*. That is, the excess returns R_i of any individual asset i can be expressed as a combination of: (i) the $N \times 1$ vector of excess returns of* all other *of risky assets $\boldsymbol{R}_{-i}$; (ii) the factor loadings on some K risky asset returns $\boldsymbol{\Phi}_i$, and we will call these K assets as the* insurance assets of asset *i; (iii) the $N \times 1$ vector of idiosyncratic risks $\boldsymbol{\varepsilon}_{-i}$ of all other N risky assets; and (iv) the idiosyncratic risk ε_i of asset i itself:*

$$R_i = \alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{F} + \varepsilon_i \tag{1a}$$

$$= \mathbf{b}_i^\top \boldsymbol{R}_{-i} + \mathbf{a}_i^\top \boldsymbol{\Phi}_i - \mathbf{b}_i^\top \boldsymbol{\varepsilon}_{-i} + \varepsilon_i. \tag{1b}$$

The $N \times 1$ vector $\mathbf{b}_i$ is dependent on the intercepts $\{\alpha_j\}_{j=1,j\neq i}^{N+1}$ and the entire factor loadings $\{\boldsymbol{\beta}_j\}_{j=1,j\neq i}^{N+1}$ of the economy and whose analytical expression is in the Appendix, and where $\mathbf{a}_i^\top := [\alpha_i, \boldsymbol{\beta}_i^\top]$.

This tells us that given *any* factor structure in the financial markets like (1a), of which its theoretical existence can always be justified via Ross (1976), the returns of a single risky asset R_i can be expressed as a linear combination $\mathbf{b}_i$ of returns of *all other* risky assets $\boldsymbol{R}_{-i}$ like (1b). The key intuition of this result is a hedging and replication argument. Suppose the researcher does *not* know the exact identifies of the factors $\boldsymbol{F} = [F_1, \ldots, F_K]^\top$ in the economy. But as long as all risky assets have an exposure to these factors, then any particular risky asset i can use some combination $\mathbf{b}_i$ of other risky stocks to hedge against risky asset i's risks. For the empirical component of our paper, this means projecting one stock's returns onto the returns of all other stocks is not a naive statistical exercise but actually has concrete microeconomic foundations.

The next result will tell us the economic content of Theorem 1.1.

Corollary 1.2. *Fix a risky asset i. Then denoting $\boldsymbol{\Phi}_i^\top := [\Phi_{0,1}, \Phi_{1,1}, \ldots, \Phi_{i,K}]$ there exists some $N \times 1$ deterministic vector $\mathbf{c}_{i,k}$ such that the k-th insurance asset return of asset i is given by,*

$$\Phi_{i,0} = \begin{cases} 1, & \text{if } \alpha_j = 0 \text{ for all } j \neq i \\ 1 - \frac{\boldsymbol{\alpha}_{-i}^\top}{||\boldsymbol{\alpha}_{-i}||_2^2}(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}), & \text{if otherwise} \end{cases} \tag{2a}$$

$$\Phi_{i,k} = F_k - \mathbf{c}_{i,k}^\top(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}), \quad k = 1, \ldots, K \tag{2b}$$

and where the form of $\mathbf{c}_{i,k}$ only depends on the factor loadings $\{\boldsymbol{\beta}_j\}_{j=1,j\neq i}^{N+1}$ and intercept $\{\alpha_j\}_{j=1,j\neq i}^{N+1}$ structure of the economy, and its analytical expression is in the Appendix. Here $||\cdot||_2$ is the Euclidean norm on $\mathbb{R}^N$, and we denote $\boldsymbol{\alpha}_{-i}^\top := [\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_{N+1}]$.

Corollary 1.2 is what motivates us to refer to those K risky assets with returns $\boldsymbol{\Phi}_i$ as *insurance assets for asset i*. From (1b), we see the following decomposition,

$$\mathbf{a}_i^\top \boldsymbol{\Phi}_i = \alpha_i \Phi_{i,0} + \boldsymbol{\beta}_i^\top \boldsymbol{\Phi}_{i,1:K}, \tag{3}$$

where $\boldsymbol{\Phi}_{i,1:K}^\top := [\Phi_{i,1}, \ldots, \Phi_{i,K}]$. Thus insurance asset of asset i can be decomposed into two sets of components: the first part $\Phi_{i,0}$ that adjusts for any mispricings, and the next K parts $\Phi_{i,k}$ adjusts for factor exposures.

The first part $\Phi_{i,0}$ from (2a) adjusts for any potential mispricings in the economy. As an important special case when there is no pricing at all in the economy [3] meaning $\alpha_i = 0$ and $\alpha_j = 0$ for all assets $j \neq i$, then $\Phi_{i,0} = 1$ from (2a) and $\alpha_i \Phi_{i,0} = 0$. If on the other hand, when the intercepts α_i, α_j's are generically

[3] It is well known that if all risky assets are mean-variance efficient then necessarily there is no mispricing in the economy. See an early discussion of this classical empirical asset pricing test in Black, Jensen, and Scholes (1972).

non-zero, then the insurance asset of asset i will explicitly hedge out any level of mispricing in the economy through $\Phi_{i,0}$, and asset i's net mispricing exposure is $\alpha_i \Phi_{i,0} = \alpha_i - \alpha_i \boldsymbol{\alpha}_{-i}^\top (\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}) / \, ||\boldsymbol{\alpha}_{-i}||_2^2$.

Secondly, the next K parts $\Phi_{i,k}$ from (2b) adjust for factor exposures. For the sake of exposition, let's suppose the K factors with returns $\boldsymbol{F}$ are tradable, even though this assumption is not needed in the proof. From the perspective of an agent who owns asset i, he is exposed to each of the $k = 1, \ldots, K$ factor returns through the factor loadings $\boldsymbol{\beta}_i^\top = [\beta_{i,1}, \ldots, \beta_{i,K}]$ of (1a). However, all the other N risky assets will also have some exposure to the k-th factor. Suppose the agent constructs an "artificial asset" that loads into the return of the k-th factor, while shorting some combination $\mathbf{c}_{i,k}$ of the factor contributions to all the other N risky assets $\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}$. The resulting artificial asset has the returns $\Phi_{i,k}$ of (2b). Hence $\Phi_{i,k}$ is precisely the residual exposure of asset i to the k-th factor, after using the returns of all other assets to "hedge" as much as possible this factor risk. The "hedging" nature of these artificial assets motivates us to call them an "insurance" for asset i. The holder of asset i is exposed to K number of factor risks, and so he will have to construct K number of these insurance assets. And since asset i has factor exposure to K number of factors, the agent will weigh $\beta_{i,k}$ loadings into the k-th insurance asset return $\Phi_{i,k}$, as in the $\boldsymbol{\beta}_i^\top \boldsymbol{\Phi}_{i,1:K}$ term of (3).

Let's rearrange (1b) and take its expectation,

$$\mathbb{E}[R_i] - \mathbf{b}_i^\top \mathbb{E}[\boldsymbol{R}_{-i}] = \mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i] - \mathbf{b}_i^\top \mathbb{E}[\boldsymbol{\varepsilon}_{-i}] + \mathbb{E}[\varepsilon_i] = \mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i], \tag{4}$$

where $\mathbb{E}[\boldsymbol{\varepsilon}_{-i}] = \mathbf{0}_N$ and $\mathbb{E}[\varepsilon_i] = 0$ because idiosyncratic risks are not priced. The overall term $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]$ in (4) is exactly the expected portfolio return into these K insurance assets. This is why we will call the term $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]$ as the *asset insurance premium for asset i.* The key objective of this paper is to empirically study (4).

The next result shows that, under weak economic and technical conditions, the asset insurance premium is non-zero.

Corollary 1.3. *The asset insurance premium of any non-redundant asset i is almost surely non-zero when some of the K factors are correlated.*

If asset i is a redundant asset, meaning it's returns are simply a linear combination of some other risky assets, then asset i's factor risks can be perfectly hedged; in this case, it is easy to show that the asset insurance premium of this redundant asset simply reduces to the factor premium for asset i. The interesting case is when asset i is *not* a perfect combination of other risky assets. The most important economic assumption in Corollary 1.3 is that there exists some correlation structure amongst the K factors. It is empirically well documented that the factors are correlated (see Bai and Ng (2006) for a discussion). The intuition of this result is similar to standard mean-variance portfolio diversification arguments. Even if the econometrician does not know the exact identifies of the K factors, but suppose the econometrician is aware that some of the factors are correlated with each other. Recall the risky assets are linear combinations of these factors. Then asset i can only beneficially use other risky assets to hedge assets against its own risk exposure if these risky assets are correlated with each other, but that is equivalent to requiring some of the factors are correlated with each other. Hence as long as there is some degree of factor correlation in the economy, there will be some benefit for asset i to construct these insurance assets, and thus there will be a non-zero return associated these insurance assets.

At this point in the theory, we have identified the existence of a quantity $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]$ that is represented as the difference in expectations as in (4). We have not pinpointed its sign and magnitude. Without using a richer theoretical setup, it is entirely an empirical question to estimate its sign and magnitude. In addition, given that the linear regression R-squared from (1a) precisely measures the amount of co-movement between asset i's return R_i with all other assets' returns $\boldsymbol{R}_{-i}$, we also call this regression R-squared the *asset synchronicity* of asset i; an asset i has lowly (highly) synchronized with all other assets if it has a low (high) R-squared.

The key contributions of this paper is to empirically show two results: *(i) an asset insurance premium exists, and is* unconditionally *positive-valued and economically large; and (ii) assets with* low (high) *asset synchronicity earn a* high (low) *asset insurance premium in the* cross-section.

In addition, the important empirical implication of (4) is that its left-hand side does *not* contain any unobservable terms, even though the right-hand side asset insurance premium term clearly depends both on the exact identification of the set of factors $\boldsymbol{F}$ and also on the exact intercepts and factor loadings $\{\mathbf{a}_j\}_{j=1}^{N+1}$. In other words, the left-hand side is econometrically a "pivotal quantity". The fact that the left-hand side quantity is a pivot is important because the asset pricing literature (e.g. Harvey et al. (2016)) has documented well above 300 known pricing factors, and it is still an open research question as to how one can shrink down this large set of factors. Theorem 1.1 tells us that we can estimate the asset insurance premium of an asset i using linear regression and portfolio sort methods, even if we do not know the exact set of the factors in the economy, nor asset i's loadings onto these factors.

The following result relates the regression R-squared to Theorem 1.1.

Corollary 1.4. *Suppose we view* (1b) *as a linear regression of* R_i *onto the set of regressors* $\boldsymbol{R}_{-i}$*, and* $\mathbf{b}_i$ *as the vector of regression coefficients. Then provided that* $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i] \neq 0$*,*

(a) The regression R*-squared decreases (increases) as* $\mathbf{a}_i^\top (\mathrm{Var}(\boldsymbol{\Phi}_i) - \mathrm{Var}(\boldsymbol{F}))\mathbf{a}_i - \sigma_\varepsilon^2$ *becomes more positive (more negative).*

(b) The regression R*-squared is decreasing in* $|\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]|$.

Corollary 1.4(a) provides another way to view of the K insurance assets of stock i from (2b). For the sake of exposition, consider the case when σ_ε^2 is negligible, and so the magnitude of the term in Corollary 1.4(a) is driven by the positive- or negative-definiteness of the matrix $\mathrm{Var}(\boldsymbol{\Phi}_i) - \mathrm{Var}(\boldsymbol{F})$. The case where this $K \times K$ matrix is *negative-definite* is when the volatility of the insurance assets of asset i is lower than the volatility of the factors themselves. This happens when the insurance assets of asset i does a good job in hedging asset i against its exposure to the K factor risks. Recall from (2b) these K insurance assets are dependent on the factor structure of all the other N risky assets. This implies these K insurance assets can only do a good job in insuring asset i against factor risks if the N other risky assets also highly co-move with asset i itself, which implies a *high* regression R-squared. Given the desirability of these K insurance assets, the holder of asset i will be willing to pay a high price for these K insurance assets, which then pushes *down* their expected returns. This explains why in Corollary 1.4(b), there is a negative relationship between the regression R-squared and the asset insurance premium $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]$. The discussion for the case when that $K \times K$ matrix is positive-definite is analogous. The above discussions still contingents upon the existence and positivity of such an asset insurance premium, which again, is entirely an empirical question we now proceed to answer.

2 Empirical hypothesis and methodology

We summarize the empirical implications of Theorem 1.1:

Empirical Hypothesis. *(i) There is microeconomic foundations for expressing the returns of a risky asset in terms of a linear combination of returns of other assets; and* *(ii) The expected difference between a given stock's return and some linear combination of other stock's returns, as in* (4)*, can be seen as an* asset insurance premium. *Under weak economic and technical conditions, the asset insurance premium is non-zero. Moreover, one does* not *need to know a priori what are the underlying factors that drive the economy to compute this difference.*

The empirical roadmap of the paper is separated into two distinct steps. By (i), the first step in our empirical methodology requires us to project a given stock's return onto the span of all other stocks' returns to get an empirical approximation of $\mathbf{b}_i$ from (1b). However, despite the microfoundations of Theorem 1.1, we shall argue that it is econometrically non-trivial execute this projection. We will apply a machine learning method to overcome a critical technical hurdle. By (ii), in the second step we will use standard portfolio sort methods from the empirical asset pricing literature to estimate the expectation $\mathbf{a}_i^\top \mathbb{E}[\mathbf{\Phi}_i]$ of (4).

2.1 Data and Projection procedure via the elastic-net

Our data source is standard. We use both the CRSP daily and monthly data from December 1974 to December 2017 with the standard filters. [4] Moreover, we only include stocks that, in the past twelve months for any given month end, there are at least 60 days of valid trading returns. This 60 days choice ensures that we do include effectively all stocks, except for the most extremely illiquid or dead stocks, so that our results are not driven only by the liquid and hence most likely large stocks. We also obtain the Fama-French data from Kenneth French's website.

[Figure 1 about here.]

The first step in the empirical test of Theorem 1.1 is to project a given stock's returns onto the returns span of all other stocks. Our empirical projection procedure is illustrated in Figure 1. For month ends $t-1 =$ December 31, 1975, January 31, 1976, ..., November 30, 2017, we use the past twelve months' worth of daily observations to project each of stock i's returns onto the returns of *all other* stocks. Unless specified otherwise, we will denote $t-1$ as the end of the projection month, and t as the one-month ahead date. Figure 2 shows the number of stocks that have at least 60 past trading days in for each given month $t-1$.

[Figure 2 about here.]

[4] We subset only for US common equities (i.e. `SHRCD` code of 10 or 11), and only those that are listed in the NYSE, AMEX or NASDAQ (i.e. `EXCHCD` code of 1, 31, 2, 32, or 3).

The return vector of stock i at month $t-1$ and the returns span of all other assets are, respectively:

$$y_{i,t-1} = \begin{pmatrix} R_{i,d_1} \\ \vdots \\ R_{i,d_{D_{i,t-1}}} \end{pmatrix}_{D_{i,t-1} \times 1} \tag{5a}$$

$$\boldsymbol{X}_{i,t-1} = \begin{pmatrix} R_{1,d_1} & \dots & R_{i-1,d_1} & R_{i+1,d_1} & \dots & R_{N_{i,t-1},d_1} \\ \vdots & & \vdots & \vdots & & \vdots \\ R_{1,d_{D_{i,t-1}}} & \dots & R_{i-1,d_{D_{i,t-1}}} & R_{i+1,d_{D_{i,t-1}}} & \dots & R_{N_{i,t-1},d_{D_{i,t-1}}} \end{pmatrix}_{D_{i,t-1} \times N_{i,t-1}} \tag{5b}$$

where $R_{i,d}$ is the daily return of stock i, $D_{i,t-1}$ is the number of trading days of stock i in the past 12 months ending at month $t-1$, and $N_{i,t-1}$ is the number stocks other than stock i. The dimensions of (2.1) is approximately 250×4000 for each stock i. There are $T = 504$ number of months from December 31, 1975 to December 31, 2017. Thus we run a total of approximately $T \times 4000 \approx 2$ million projections in this paper.

In this paper we will use the *elastic-net estimator* developed by Zou and Hastie (2005) to empirically project a given stock's return onto the returns span of all other stocks. But if we are interested in estimating *linear* relationships as in Theorem 1.1, why do we not use the workhorse *ordinary least squares* (OLS) estimator? The design matrix of returns to evaluate our empirical hypothesis is necessarily a $T \times N$ matrix, where $T \approx 250$ is the number of days, and $N \approx 4000$ is the total number of traded stocks. This is a case where $T \ll N$. This means the $N \times N$ matrix $\boldsymbol{X}_{i,t-1}^{\top}\boldsymbol{X}_{i,t-1}$ is *not* full rank. The OLS estimator is thus necessarily *not* well defined. In contrast, the elastic-net is a machine learning method that explicitly allows for "wide" $T \ll N$ regressors. Section A.2 in the Appendix gives a brief technical introduction to this method. Moreover, Section A.2 discusses how the elastic-net can be seen as projecting a constrained *linear* relationship between the dependent and the independent variables. Out of a myriad of machine learning methods, why did we choose the elastic-net estimator to test our empirical implication? We will have some further remarks on this matter in Section 2.3.1 after we discuss our portfolio construction method.

We make clear that we are only using the elastic-net as a projection method, and we do *not* use it for statistical inference. The statistical inference claims are on the expected returns of the asset insurance premium, which we discuss beginning in Section 2.3. As a result, despite the large number of projections that we run at this stage, we do not suffer from the multiple hypothesis testing problem that has been discussed in the recent empirical asset pricing literature by Harvey et al. (2016).

It is evident that we are using overlapping time data. Overlapping returns data in traditional empirical asset pricing papers raises several technical inference issues revolving around autocorrelations (see Hansen and Hodrick (1980) and a recent discussion by Hedegaard and Hodrick (2016)). In contrast, an implication of using "wide" regressors in our setting is that we actually gain ≈ 4000 new cross-sectional data points for each time advance. As a result, simply because of the sheer amount of new entering cross sectional data, it is not necessarily true that the estimated coefficient ending at months $t-2$ and $t-1$ would be quantitatively similar even if they only differ by one month step.

We denote the elastic-net estimated coefficient vector of stock i at month $t-1$ as $\hat{\beta}_{i,t-1} \in \mathbb{R}^{N_{i,t-1}}$, and the resulting *coefficient of determination ("R-squared")* as $\mathsf{R}^2{}_{i,t-1}$. As we shall further justify in Section 4, we will also call $\mathsf{R}^2{}_{i,t-1}$ as the *asset synchronicity* of stock i. Note and recall that the conventional definition of R^2 is,

$$\mathsf{R}^2_{i,t-1} := 1 - \frac{\sum_{s=1}^{D_{i,t-1}} (R_{i,d_s} - \hat{y}_{i,t-1})^2}{\sum_{s=1}^{D_{i,t-1}} (R_{i,d_s} - \hat{\mu}_{i,t-1})^2}, \tag{6}$$

where $\hat{y}_{i,t-1} := \boldsymbol{X}_{i,t-1}\hat{\beta}_{i,t-1}$ is the fitted value, and $\hat{\mu}_{i,t-1}$ is the sample mean; they are explicitly given by

$$\begin{aligned} \hat{y}_{i,t-1} &= \boldsymbol{X}_{i,t-1}\hat{\beta}_{i,t-1}, \\ \hat{\mu}_{i,t-1} &= \frac{1}{D_{i,t-1}} \sum_{s=1}^{D_{i,t-1}} R_{i,d_s}. \end{aligned}$$

Table 1 shows the summary statistics of the estimated elastic-net coefficients. What is striking is that despite the large number of stocks that are present in the cross-section at any point in time, the elastic-net only selects (i.e. has non-zero coefficient loadings) a very small subset of stocks to fit any given stock i. For instance, stocks with the lowest (highest) R^2's have an average of 0.505 (79.032) non-zero entries in their high-dimensional estimated elastic-net coefficient vector. This is the empirical realization of the "sparsity effect" of the elastic-net estimator. Moreover, high R^2's of actual stock i is positively associated with the count of non-zero elements in the estimated elastic-net coefficient vector. [5] In addition, the dispersion summary statistics (i.e. standard deviation, median and percentiles) show that the estimated coefficients are

[5] While there is some resemblance, we caution that the positive association of R^2 and number of non-zero elements in the estimated elastic-net coefficient vector is different from the OLS case where including more regressors necessarily raises R^2. In our elastic-net application, the number of regressors $N_{i,t-1}$ remains *fixed*, and it is an estimation outcome that only few of them have non-zero coefficient loadings.

fairly stable and concentrated. This is the empirical realization of the "grouping property" of the elastic-net estimator. We will elaborate further on this grouping property in Section 5.

[Table 1 about here.]

Table 2 shows the summary statistics of characteristics of actual stocks as sorted by their elastic-net R^2. We make several noteworthy observations. High elastic-net R^2 appears to be positively associated with stocks of: (i) low book-to-market ratio and (ii) high dollar trading volume. While the highest elastic-net R^2 stocks are associated with the largest market capitalizations, we do not observe a monotonic pattern in this relationship with the rest of the other R^2 bins. There is little evidence that R^2's are associated with total volatility. The bin sorts suggest that there is a wide heterogeneity the explanation power of all other stocks on one given stock: the lowest bin only has -1.2% mean elastic-net R^2, while the highest bin attains a mean of 60.6%.

[Table 2 about here.]

2.2 The "ghost portfolio" of asset i

In this section, we will track three types of returns for each stock, which will then lead to three different definitions of portfolio returns. Figure 3 shows an illustration. We will be technically explicit in making clear our definitions and constructions in this section to avoid any ambiguities.

We discuss our portfolio construction procedure and introduce our key idea of the *ghost of stock* i. For each month end $t-1$, we collect the projected coefficient $\hat{\beta}_{i,t-1}$ and the $\mathsf{R}^2{}_{i,t-1}$ for each stock $i = 1, \ldots, N_{i,t-1}$. First, we track stock i's *one-month ahead* return, and we call this the return of the *actual stock* i at month t,

$$R_{i,t}^{\text{Act}} \equiv R_{i,t}, \tag{7}$$

Here, we consider the months $t =$ January 31, 1976, February 29, 1976, $\ldots$, December 31, 2017.

Next, we introduce the key idea of this paper. Let $\boldsymbol{R}_{-i,t} := [R_{1,t}, \ldots, R_{i-1,t}, R_{i+1,t}, \ldots, R_{N_{i,t-1},t}]^\top$ be the vector of one-month ahead returns from month $t-1$ to month t for all stocks except stock i. We will treat the estimated coefficients $\hat{\beta}_{i,t-1}$ as investment weights into each of the $N_{i,t-1}$ number of stocks. Since investment weights must sum to one, we will place the remainder of the weights $\mathbf{1}^\top \hat{\beta}_{i,t-1}$ into the risk free asset with returns $r_{f,t}$, and where here $\mathbf{1}$ is a vector of ones of conformable dimensions. We define the *ghost (portfolio)* of stock i at month t as,

$$R_{i,t}^{\text{Gho}} := (1 - \mathbf{1}^\top \hat{\beta}_{i,t-1}) r_{f,t} + \boldsymbol{R}_{-i,t}^\top \hat{\beta}_{i,t-1}. \tag{8}$$

We coin the terminology "ghost" because as inspired by the conventional picture of a projected vector like Figure 3, the "ghost" exists in the in the projected shadow of its true actual self. [6] Breeden et al. (1989) and Lamont (2001) use analogous methods to (8) to construct a mimicking factor out of tradable base assets. Ang et al. (2006) use also an analogous method to construct a factor mimicking aggregate volatility risk. While our approach in (8) seems identical to these previous methods, we stress that the dimensionality of the regressors are substantially different. The number of base assets in those aforementioned methods are small, usually in the range of five to ten. In contrast, our base assets are effectively every other stock other the actual stock i itself, which number in the thousands. Most importantly, the ghost portfolio (8) will proxy for $\mathbf{b}_i^\top \boldsymbol{R}_{-i}$ in Theorem 1.1.

Third and finally, we track the *long-short* return of the actual stock i against its ghost,

$$R_{i,t}^{\text{Ls}} := R_{i,t}^{\text{Act}} - R_{i,t}^{\text{Gho}}. \tag{9}$$

This long-short return (9) will proxy for the difference $R_i - \mathbf{b}_i^\top \boldsymbol{R}_{-i}$ in Theorem 1.1. We will use conventional portfolio sort methods of the empirical asset pricing literature to estimate the expectation $\mathbb{E}[R_i] - \mathbf{b}_i^\top \mathbb{E}[\boldsymbol{R}_{-i}]$, which then is equal to the predicted asset insurance premium $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]$ of stock i.

[Figure 3 about here.]

2.3 Portfolio sort and construction

Having defined three types of returns for stock i, we now proceed to construct portfolios. We use Corollary 1.4 as guidance. Corollary 1.4 states that if an asset insurance premium exists and is positive, then one should find a negative relationship between the regression R-squared and asset insurance premium $\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]$. Thus

[6] More whimsically, we drew some inspirations from Charles Dickens' *A Christmas Carol*: "I have endeavoured in this Ghostly little book, to raise the Ghost of an Idea, which shall not put my readers out of humour with themselves, with each other, with the season, or with me."

at the end of the month $t-1$, we will sort each stock i by its $\mathsf{R}^2{}_{i,t-1}$ into *decile* bins. Let B^k_{t-1} be the set of stocks in the k-th bin at month $t-1$. We organize the bins in ascending order, so bin $k=1$ (labeled "Lo") consists of stocks with the lowest R^2's, and bin $k=10$ (labeled "Hi") consists of stocks with the highest R^2's.

The definitions of equal-weighted and value-weighted portfolio *excess returns* of bin $k=1,\ldots,10$ of the *actual* stocks are standard:

$$\bar{R}^k_{\mathrm{ActEql},t} := \frac{1}{\left|B^k_{t-1}\right|} \sum_{i \in B^k_{t-1}} (R^{\mathrm{Act}}_{i,t} - r_{f,t}), \tag{10a}$$

$$\bar{R}^k_{\mathrm{ActVal},t} := \sum_{i \in B^k_{t-1}} w^k_{i,t-1}(R^{\mathrm{Act}}_{i,t} - r_{f,t}), \tag{10b}$$

where $w^k_{i,t-1}$ is the relative market capitalization of stock i in the k-th bin at month $t-1$, and where we denote $|B|$ as the cardinality of a set B.

While the definitions of equal-weighted and value-weighted portfolios of the actual stocks are conventional, we need to apply some discretion in defining the weighting mechanism for the ghosts. In this paper, we will view the ghost of stock i to have the same weight as its actual counterpart. This means we define the equal-weighted and value-weighted portfolio *excess returns* of bin k of the *ghosts* as,

$$\bar{R}^k_{\mathrm{GhoEql},t} := \frac{1}{\left|B^k_{t-1}\right|} \sum_{i \in B^k_{t-1}} (R^{\mathrm{Gho}}_{i,t} - r_{f,t}), \tag{11a}$$

$$\bar{R}^k_{\mathrm{GhoVal},t} := \sum_{i \in B^k_{t-1}} w^k_{i,t-1}(R^{\mathrm{Gho}}_{i,t} - r_{f,t}). \tag{11b}$$

Note again that we assign weight $w^k_{i,t-1}$ of the actual stock i to its ghost counterpart.

Finally, the construction of the equal-weighted and value-weighted portfolios of the long-short of the actual versus ghost stocks is now immediate. We define the equal-weighted and value-weighted portfolio *excess returns* of bin k of the *long-short actual versus ghost* as,

$$\bar{R}^k_{\mathrm{LsEql},t} := \bar{R}^k_{\mathrm{ActEql},t} - \bar{R}^k_{\mathrm{GhoEql},t} = \frac{1}{\left|B^k_{t-1}\right|} \sum_{i \in B^k_{t-1}} R^{\mathrm{Ls}}_{i,t}, \tag{12a}$$

$$\bar{R}^k_{\mathrm{LsVal},t} := \bar{R}^k_{\mathrm{ActVal},t} - \bar{R}^k_{\mathrm{GhoVal},t} = \sum_{i \in B^k_{t-1}} w^k_{i,t-1} R^{\mathrm{Ls}}_{i,t}. \tag{12b}$$

Unlike the definitions of excess returns of the actual stocks (10) and of the ghost stocks (11), the definition of excess returns in the actuals minus ghosts (12) does not involve subtracting off a risk free rate. Conventionally, longing a risky asset and shorting a risk free asset can be viewed as the "risk premium" of holding the risky asset. However, we can also view this long risky asset, short risk free asset as a zero-cost trading strategy. We primarily take the latter view in the definition of (12). Indeed, if we were to subtract off another risk free rate in (12), the resulting portfolio strategy would require an investor to take a long position in the actual stocks, short the ghost stocks, and short the risk free asset, hence resulting the investor in a "naked" costly position in one of the short legs. Finally, a "premium" interpretation as motivated from Theorem 1.1 also becomes challenging when the round trip trading strategy is actually costly and not zero-cost.

2.3.1 Parsimony and the economic rationale of the elastic-net

We can now justify why we choose to implement the elastic-net estimator in this paper out of the myriad of machine learning methods. Without a doubt, there are more advanced machine learning and econometric methods that can increase the in-sample fit, and thereby boost the in-sample R^2. A recent extensive study by Gu, Kelly, and Xiu (2018) show that many possible machine learning methods can achieve high in-sample and out-of-sample R^2's. However, the issue with these "black box" methods is that regardless of their in-sample or even out-of-sample performance, their estimated model parameters often do not have a transparent link to the regressors themselves. In contrast, although the elastic-net estimator is inherently non-linear, its estimated coefficients can be applied in linearly back to the regressors. This parsimonious nature of the elastic-net allows us to explicitly and linearly construct the ghost portfolios as in (8).

We emphasize that the elastic-net machine learning method is simply a tool to test the prediction of Theorem 1.1 via the construction of the ghosts (8). We have no intentions to conduct statistical inference on the estimated elastic-net coefficients. Our statistical inference claims are still based upon well understood portfolio sort methods in the empirical asset pricing literature as outlined in Section 2.3.

The consideration of the ghosts is what drives us to prefer the elastic-net over its close cousin, the LASSO. The LASSO enjoys a "sparsity property", whereby the number of estimated coefficients tend to be small, even though the set of regressors could be large. Sparsity means there are only a handful of stocks that have to considered in order to construct the ghost of stock i. This is important because if the number of stocks needed to construct the ghost of stock i is large, then whatever statistical results we claim to find may be economically infeasible due to trading costs and other market frictions. Unfortunately, if a group of regressors are highly correlated with each other then LASSO has a tendency to only select one regressor at effectively at random. This makes for a poor portfolio construction of the ghosts for numerous reasons, and loss of potential diversification is an obvious one. The elastic-net remedies this problem by inheriting the grouping property of the ridge estimator. In all, this means the elastic-net is a good candidate for our consideration of the ghosts because: (i) it can fit the data (i.e. from the least squares property of the OLS); (ii) it can linearly apply its estimated coefficients over the regressors; (iii) it has a sparsity property (i.e. from LASSO); and (iv) it has a grouping property (i.e. from ridge). In Section A.3.2 we have further technical discussions on the importance of this parsimony.

3 Main Results I: Unconditional average asset insurance premium and in the Cross-section of asset synchronicity

The univariate portfolio sorts of Table 3 shows the main empirical results of this paper. Let's discuss the actual returns of stocks in the first column of Table 3. In the cross-section with equal-weight [value-weight] portfolios, stocks with the *lowest* elastic-net R^2 earn an actual return of 1.440% (t-stat 5.042) [0.969% (t-stat 5.304)] per month. Stocks with the *highest* elastic-net R^2 earn an actual return of 0.786% (t-stat 2.827) [0.628% (t-stat 2.677)]. The *difference* in actual returns between stocks with the highest and lowest elastic-net R^2 is -0.653% (t-stat -4.246) [-0.341% (t-stat -2.174)]. In other words, lowly synchronized stocks have an actual return that is higher than the returns of highly synchronized stocks.

By conventional standards in the empirical asset pricing literature, the above results strongly suggest that R^2 is a potential priced factor in the cross-section. However, guided by Theorem 1.1, we are not just interested in testing the cross-sectional *difference* in the actual stocks. Rather, we are interested in empirically testing for the presence of an asset insurance premium of (4). Let's discuss the empirical results of the ghost portfolios in the second column of Table 3. First and foremost is the absence of statistically significant results across all the R^2 bins. The point estimates in equal-weighted ghosts are monotonically increasing, from -0.000% (t-stat -0.204) per month in the *lowest* R^2 bin, to 0.416% (t-stat 1.670) in the *highest* R^2 bin, even though their *difference* 0.417% (t-stat 1.674) is statistically insignificant. The same can be said with value-weighted ghosts, where the point estimates are -0.001% (t-stat -0.418) in the *lowest* R^2

bin, and almost monotonically increases to 0.325% (t-stat 1.617) in the *highest* bin, again even though the *difference* 0.326% (t-stat 1.623) is statistically insignificant. Despite the lack of statistical significance, by construction of the ghost portfolio (8), the monotonic increase in the point estimates can only be explained because stocks with low R^2 are associated with almost full loadings into the risk free asset, while stocks with high R^2 are associated with high loadings into risky assets; this is confirmed in Table 1.

Finally, we come to the main highlight and prediction of our paper: the results of the actuals minus the ghosts in the third column of Table 3, which proxy for the asset insurance premium $\mathbf{a}_i^\top \mathbb{E}[\mathbf{\Phi}_i]$ in Theorem 1.1. For each R^2 bin, we see that the returns of the actuals minus ghosts are statistically significant, positive, and also monotonically decreasing in R^2. With equal-weights, the actuals minus ghosts have an average return of 1.440% (t-stat 5.027) per month for stocks with the *lowest* R^2's, it is 0.371% (t-stat 3.245) with the *highest* R^2's, and the *difference* of -1.069% (t-stat -3.904) is highly statistically significant. Value-weight results are analogous, where the average returns of actuals minus ghosts is 0.976% (t-stat 5.081) in the *lowest* R^2 bin, 0.305% in the *highest* R^2 bin, while the *difference* of -0.671% (t-stat -3.536) is highly statistically significant.

In addition to observing a cross-sectional difference in asset insurance premium across asset synchronicity, an asset insurance premium also exists for an average risky asset. The "Avg" equal-weight portfolio in Table 3 takes the simple average of returns of all of the ten $k = \text{'Lo'}, 2, \dots, \text{'Hi'}$ equal-weight portfolios. Likewise, the "Avg" value-weight portfolio in Table 3 takes the simple average of returns of all of the ten $k = \text{'Lo'}, 2, \dots, \text{'Hi'}$ value-weight portfolios. With equal-weight [value-weight] portfolios, the average asset insurance premium is 0.793% (t-stat 3.942) [0.575% (t-stat 4.333)].

In all, these empirical results justify and sharpen the core predictions of Theorem 1.1: *(i) an unconditional asset insurance premium exists and is positive for the* average *stock; and (ii) stocks with low (high) asset synchronicity earn a high (low) asset insurance premium in the* cross-section.

[Table 3 about here.]

3.1 Correlation between actuals and ghosts

A potential concern is that the statistically significant results in the actuals minus ghosts follow trivially from the statistically results of the actuals. This is not necessarily the case. Firstly, taking one sequence of statistically significant (from zero) $\{Y_i\}$ variables and subtracting another sequence of statistically insignificant variables $\{X_i\}$ does not necessarily lead to a sequence of statistically significant variables. The variance of the difference is $\text{Var}(Y_i - X_i) = \text{Var}(Y_i) + \text{Var}(X_i) - 2\text{Corr}(Y_i, X_i)\text{Std}(Y_i)\text{Std}(Y_i)$. Even in the extreme case that X_i is mean zero and independent of Y_i, its variance $\text{Var}(X_i)$ will still positively contribute to the overall variance $\text{Var}(Y_i - X_i)$, and thereby makes inferring whether $Y_i - X_i$ is statistically significantly different from zero even more challenging. The interesting case where subtracting X_i can contribute a smaller term to $\text{Var}(Y_i - X_i)$ is when $\text{Corr}(Y_i, X_i)$ is positive and high — this means $\{Y_i\}$ and $\{X_i\}$ could have an economic hedging relationship with each other.

Motivated by this possibility, Table 4 investigates the pairwise correlations between the portfolio of actual stocks against the portfolio of ghost stocks per R^2 bin. The diagonal entries of Table 4 are most important. Stocks with low (high) R^2's have low (high) correlations between their actual returns versus their ghosts. In equal-weights, the correlation between the one-month ahead returns of the actual stocks and their ghost counterparts with the *lowest* R^2 is only -9.6%; in contrast, this correlation is 92.1% in the *highest* R^2 bin. In value-weights, the correlation between the one-month ahead returns of the actual stocks and their ghost counterparts with the *lowest* R^2 is only -7.7%, but it is 87.7% in the *highest* R^2 bin. Economically, this means stocks that are highly synchronized with other stocks can use the ghosts as an effective hedging instrument, even if ghosts have extremely low returns. However, stocks that are lowly synchronized with each other will not find other stocks to hedge against itself. This is precisely the argument that we made in constructing the returns $\mathbf{\Phi}_i$ of the insurance asset for asset i from Theorem 1.1.

[Table 4 about here.]

3.2 Risk adjusted returns on post-formation portfolios

The main results Table 3 do not control for priced risk factors. We now show the asset insurance premium results still hold even after adjusting for the Fama-French three factors (Fama and French (1992, 1993)) and Fama-French five factors (Fama and French (2015)). For notational brevity, we will henceforth refer to the Fama-French z factors as FFz, where $z = 3, 5$.

After we have constructed the elastic-net R^2 decile portfolios according to Section 2.2 for all months

$t = 1, \dots, T$, we consider the following two time series factor model regressions on each bin $k = 1, \dots, 10$:

$$r_t^k = \alpha^k + \beta_{\text{MktRF}}^k \text{MktRF}_t + \beta_{\text{SMB}}^k \text{SMB}_t + \beta_{\text{HML}}^k \text{HML}_t + \varepsilon_t^k, \tag{13}$$

and

$$r_t^k = \alpha^k + \beta_{\text{MktRF}}^k \text{MktRF}_t + \beta_{\text{SMB}}^k \text{SMB}_t + \beta_{\text{HML}}^k \text{HML}_t + \beta_{\text{CMA}}^k \text{CMA}_t + \beta_{\text{RMW}}^k \text{RMW}_t + \varepsilon_t^k. \tag{14}$$

The regressors are well-known: "MktRF" is the market factor, "SMB" is the size factor, "HML" is the value factor, "CMA" is the investment factor, and "RMW" is the profitability factor.

The response variable of (13) and (14) depends on the return and portfolio weighting type. For actual stocks and the ghost stocks, we take the excess returns that subtract off the risk free rate, so $r_t^k \in \{\bar{R}_{\text{ActEql},t}^k, \bar{R}_{\text{ActVal},t}^k\}$, and $r_t^k \in \{\bar{R}_{\text{GhoEql},t}^k, \bar{R}_{\text{GhoVal},t}^k\}$, of (10) and (11) respectively. But for the long-short returns of actuals against its ghosts, we do not subtract off the risk free rate for reasons discussed in Section 2.2, and so $r_t^k \in \{\bar{R}_{\text{LsEql},t}^k, \bar{R}_{\text{LsVal},t}^k\}$ of (12).

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

Table 5 (equal-weight) and Table 6 (value-weight) show the FF3 regression results of (13), while Table 7 (equal-weight) and Table 8 (value-weight) show the FF5 regression results of (14). The results confirm the findings of our univariate portfolio sort results of Table 3 after adjusting for factor risks. In each of the elastic-net R^2 bins, we see that the return from a long position in the actual stocks portfolio, and short position in the its ghost counterpart portfolio generates economically large and statistically significant alphas. In the FF3 regression with equal-weights, this long-short alpha almost monotonically decreases from 1.076% (t-stat 4.503) per month in the *lowest* bin, to 0.407% (t-stat 3.475) in the *highest* bin, and their difference is -0.669% (t-stat -2.931). The FF3 regression with value-weights tell a similar story, where the long-short alpha almost monotonically decreases from 0.920% (t-stat 4.857) per month in the *lowest* bin, to 0.409% (t-stat 3.414) in the *highest* bin, with their *difference* being -0.511% (t-stat -2.904). The FF5 regressions tell qualitatively the same story but with even stronger statistical evidence. The FF5 alphas of this long-short strategy decrease almost monotonically from 1.179% (t-stat 4.758) per month in the *lowest* bin, to 0.451% (t-stat 3.834) in the *highest* bin, with their *difference* being -0.728% (t-stat -3.056). Likewise, the FF5 alphas of this long-short strategy with value-weights decrease almost monotonically from 1.030% (t-stat 5.215) per month in the *lowest* bin, to 0.428% (t-stat 3.425) in the *highest* bin, with their *difference* being -0.602% (t-stat -3.228). In all, the cross-section of asset insurance premium still persists even after controlling for well-known risk factors.

4 Main Results II: Asset synchronicity and macroeconomic countercyclicality

There are two main empirical findings of Section 3: (i) unconditionally there exists a positive asset insurance premium for an *average* stock; and (ii) lowly (highly) synchronized stocks earn a high (low) asset insurance premium in the *cross-section*. Here we provide a macroeconomic justification for these two empirical findings.

[Figure 4 about here.]

Figure 4 plots the time series of the cross-sectional R^2 averages of the elastic-net. A cursory observation suggests that the levels of the cross-sectional R^2 averages could be *countercyclical*. Asset synchronicity spikes up considerably during recessions and also during times of financial distress (e.g. 1989-1991 US savings and loans crisis, 1998 Russian financial crisis, 2008-2009 Great Recession, Greek government debt crisis 2011-2012, and others). Asset synchronicity is very low during the decade long economic expansion of the 1990's. We formally investigate whether asset synchronicity could be related to macroeconomic fundamentals.

[Table 9 about here.]

By arguments from standard consumption-based asset pricing theory (see Campbell (2003) for a survey), a variable is only relevant for a representative agent in the economy if the variable has non-zero correlation with the pricing kernel. In particular, results from those theories suggest that there should only be a positive asset insurance premium only if the long-short portfolio returns are positively correlated with consumption growth. The main results of Section 3 already show that asset insurance premium and asset synchronicity are negatively related to each other. Now let's examine the relationship between asset synchronicity and consumption. Table 9 regresses the shocks to cross-sectional R^2 averages onto shocks of macroeconomic and financial variables. Indeed, we find that shocks to cross-sectional R^2 averages are negatively driven by *shocks to consumption*; a 1% change in monthly consumption is associated with a -.408% (t-stat -2.568) monthly difference in average asset synchronicity. Putting these two observations together, this suggests shocks to consumption and the asset insurance premium for an average risky asset should be positively correlated with each other. The third columns of Table 10 (equal-weight) and Table 11 (value-weight) regress the average asset insurance premium onto various macroeconomic and financial regressors. With equal-weights [value-weights] an 1% change in monthly consumption is associated with a 2.173% (t-stat 2.179) [1.853% (t-stat 2.366)] monthly change in the *unconditional average* asset insurance premium. Thus we have good evidence that an average asset insurance premium exists because of the positive empirical relationship between asset synchronicity and consumption shocks.

[Table 10 about here.]

[Table 11 about here.]

What remains is to explain the *cross-section* of asset insurance premium across lowly and highly synchronized stocks. Following the above macroeconomic arguments, our observed cross-sectional difference in the asset insurance premium across asset synchronicity is only possible when consumption shock disproportionately affect the returns of lowly and highly synchronized stocks. The third columns of Table 12 (equal-weight) and Table 13 (value-weight) regress the cross-sectional asset insurance premium onto a set of macroeconomic and financial regressors. With equal-weights [value-weights], we see that a 1% change in monthly consumption is associated with a -4.792% (t-stat -2.633) [-1.725% (t-stat -1.907)] monthly change in the cross-sectional asset insurance premium between the highly and lowly synchronized stocks. In other words, lowly synchronized stocks are more sensitive to consumption shocks than highly synchronized stocks, which justifies why lowly synchronized stocks warrant a higher asset insurance premium than highly synchronized stocks.

[Table 12 about here.]

[Table 13 about here.]

4.1 Next best alternative investment and the elastic-net estimator

We have established that asset synchronicity is negatively related to macroeconomic fundamentals, and that high asset synchronicity negatively affects asset returns. What then induces a risk averse agent to hold risky assets in light of this systematic risk? In the conventional theoretical and empirical asset pricing literature, the term $R_i - r_f$ is coined as a *risk premium* (or more precisely *equity premium* in our context) because its expected value is the amount of compensation that would induce a risk averse agent to prefer a risky gamble over the next best alternative, which is presumably a riskless gamble. However, there is no reason why a risk averse agent is effectively coerced to only choose between a risky choice and an risk less alternative. Suppose the agent is presented with a third option: a "less risky" gamble R_i^{Gho} that resembles the original risky gamble R_i. Then faced with these three choices, the next best alternative to the risky gamble R_i for the risk averse agent need not be the riskless choice, but rather the "next best less risky" gamble R_i^{Gho}. The difference $R_i - R_i^{\text{Gho}}$ is the *asset insurance premium* for inducing the risk averse agent to prefer a higher risk gamble over a closely matched lower risk gamble. This argument is made concrete by the choice of loadings $\mathbf{b}_i$ on all the other risky assets in Theorem 1.1

We can now further understand why we find an asset insurance premium in the cross-section of asset synchronicity using the elastic-net approach. According to the above argument, we should expect the "next best" alternative to stocks that are lowly synchronized with other stocks to simply be the risk free asset itself. In contrast, the "next best" alternative to stocks that are highly synchronized with other stocks should be a portfolio of other risky stocks. Figure 5 shows the time series of the equity-only component in the ghost portfolio across the elastic-net R^2 sorted bins, and this figure confirms our above arguments. Figure 5 shows the ghosts for the stocks in the lowest elastic-net R^2 bin are effectively just the risk free asset for all time, while the ghost for the stocks in the highest elastic-net R^2 bin are composed of all risky assets. Next, compare Figure 4 against Figure 5. From the countercyclicality of asset synchronicity, the decade long economic expansion of the 1990's is exactly when asset synchronicity is at its lowest levels, and reaches its highest levels during the 2008-2009 financial crisis. During times when overall asset synchronicity is low (high), our argument above anticipates that the next best alternative investment should effectively be just the risk free asset (risky assets) for the average risky asset; the subplot labeled as “Overall average” Figure 5 empirically confirms this argument.

[Figure 5 about here.]

The ghost portfolio achieves this economically sound “next best alternative” construction because of the grouping effect of the elastic-net estimator. Recall Table 1. In the lowest elastic-net R^2 bin, the average proportion of wealth allocated to other risky stocks is 0.00%, and its interquartile range is numerically undetectable at three significant decimal figures, $0.00\% - 0.00\% = 0.00\%$. This is coherent with our above argument whereby the next best alternative investment to a stock that least synchronized with every other stock in the market is simply the risk free asset. In the highest elastic-net R^2 bin, the average proportion is 53.1%, and the associated interquartile range is 77.3% - 26.1% = 51.2%. In Section 5, we will show that an analogous approach to constructing ghost portfolios using the Fama-French factors as trading assets will yield a far higher interquartile range. These tight interquartile ranges of the elastic-net method imply that we can be quite confident that ghost portfolios within each elastic-net R^2 bin are fairly homogeneous in terms of the proportion of wealth that are allocated to risky assets.

In all, our results here relate to a recent literature that links comovement risk (e.g. Adrian and Brunnermeier (2016)) and flight-to-safety (e.g. Longstaff (2004) and Beber et al. (2008)) episodes during times of crisis. As this literature has documented the overall effects of the market and also on liquidity on these crises-related variables, we take special care to control for financial regressors such as the market factor and the Pástor and Stambaugh (2003) liquidity factor in all of the tables in this section. The results in this section persist even after controlling for these two financial regressors.

5 Comparing against the ghosts of FF3 and FF5

Using the elastic-net, the paper’s main results in Section 3 show that asset synchronicity and asset insurance premium are negatively related to each other as shown in Theorem 1.1. It is very natural to ask the follow-up question: how do the results out of elastic-net R^2’s compare against the R^2’s out of benchmark models such as the Fama-French three factor and five factor models? This comparison is particularly important because if the results out of the R^2's out of the Fama-French models are similar, then it invalidates both the need to use a complex machine learning method.

We revisit the discussions of Section 2 in this new context. We use the same return vector $y_{i,t-1}$ for each stock i and each months $t-1$ as in (5a). Instead of using the returns span of all other assets as in (2.1), we project $y_{i,t-1}$ onto the returns span of the Fama-French three factors (Fama and French (1992, 1993)), and that of the Fama-French five factors (Fama and French (2015)), henceforth "FF3" and "FF5",

$$\boldsymbol{X}_{\mathrm{FF3},t-1} = \begin{pmatrix} \mathrm{MktRF}_{d_1} & \mathrm{SMB}_{d_1} & \mathrm{HML}_{d_1} \\ \vdots & \vdots & \vdots \\ \mathrm{MktRF}_{D_{i,t-1}} & \mathrm{SMB}_{D_{i,t-1}} & \mathrm{HML}_{D_{i,t-1}} \end{pmatrix}, \tag{15a}$$

$$\boldsymbol{X}_{\mathrm{FF5},t-1} = \begin{pmatrix} \mathrm{MktRF}_{d_1} & \mathrm{SMB}_{d_1} & \mathrm{HML}_{d_1} & \mathrm{CMA}_{d_1} & \mathrm{RMW}_{d_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{MktRF}_{D_{i,t-1}} & \mathrm{SMB}_{D_{i,t-1}} & \mathrm{HML}_{D_{i,t-1}} & \mathrm{CMA}_{D_{i,t-1}} & \mathrm{RMW}_{D_{i,t-1}} \end{pmatrix}. \tag{15b}$$

Relating back to Theorem 1.1, that means we regard the returns of the other assets $\boldsymbol{R}_{-i}$ as only the returns of the three and five Fama-French tradable "portfolios". It is important to emphasize that in the context of (15), we do *not* view them as priced factors, but rather we view them merely as other tradable portfolios for asset i. [7] For all months $t-1$ and all stocks i, we use the OLS estimator to project $y_{i,t-1}$ onto $\boldsymbol{X}_{\mathrm{FF}z,t-1}$ for $z = 3, 5$, and we collect the estimated OLS coefficient vector $\hat{\beta}_{\mathrm{FF}z,i,t-1}$, and their $\mathsf{R}^2{}_{\mathrm{FF}z,i,t-1}$.

We construct the ghosts of stock i using the FFz factors analogous to that of Section 2.2. The actual return of stock i remains the same as in (7). We obtain the ghost of stock i now with the one-month ahead FFz factor returns:

$$R_{i,t}^{\mathrm{Gho,FF}z} := (1 - \mathbf{1}^\top \hat{\beta}_{\mathrm{FF}z,i,t-1}) r_{f,t} + \boldsymbol{R}_{\mathrm{FF}z,t}^\top \hat{\beta}_{\mathrm{FF}z,i,t-1}, \tag{16}$$

where $\boldsymbol{R}_{\mathrm{FF}z,t}$ is a vector of month t returns of the FFz factors. The long-short return of actual stock i against its ghost is analogous to (9), where

$$R_{i,t}^{\mathrm{Ls,FF}z} := R_{i,t}^{\mathrm{Act}} - R_{i,t}^{\mathrm{Gho,FF}z}. \tag{17}$$

[7] In principle, we could directly use the Fama-French portfolios in the analysis here. For instance, rather than inserting the HML factor as a tradable portfolio into $\boldsymbol{R}_{-i}$, we could use the two constituent portfolios (i.e. $\frac{1}{2}$(Small Value + Big Value) and $\frac{1}{2}$(Small Growth + Big Growth) that build up to the HML factor. For simplicity, we will just view a factor as a single tradable portfolio.

Symmetric to Section 2.2, we sort stocks into their FFz R^2 decile bins and form equal- and value-weighted portfolios. Note that in FFz, the estimated coefficient vector $\hat{\beta}_{\mathrm{FF}z,i,t-1}$ of stock i at month $t-1$ is always $z \times 1$-dimensional and where $z = 3, 5$. In contrast, the dimensionality of the elastic-net estimated coefficient vector $\hat{\beta}_{i,t-1}$ is $N_{i,t-1} \times 1$, meaning the dimensionality explicitly depends on the stock and also in time.

We first examine the relationship between elastic-net R^2's versus FFz R^2's. Table 14 shows the unconditional correlation between the R^2's of the elastic-net, FF3 OLS and FF5 OLS regression estimation methods. Unsurprisingly, the R^2's from FF3 and FF5 are effectively perfectly positively correlated at 0.995. We also observe high correlations of the elastic-net R^2 with that of the FFz R^2, at 0.815 with FF3, and 0.815 with FF5. On the one hand, this suggests the explanatory power of the elastic-net closely tracks well established factor model models in the empirical asset pricing literature — despite that we did not impose any factor structure in the elastic-net estimator, and indeed the FFz factors were never in the set of regressors (5). On the other hand, the high correlations between elastic-net R^2's and FFz R^2's raise the question of whether one can obtain the same result as in Table 3 but just construct portfolios by sorting FFz R^2's.

[Table 14 about here.]

5.1 No asset insurance premium in the cross-section of FF3- and FF5-based asset synchronicities

Table 15 (FF3) and Table 16 (FF5) show the counterpart portfolio sort by FFz R^2, and these are direct counterparts to Table 3. We *fail* to find evidence to suggest that sorting by FF3 and FF5 R^2 yields a cross-sectional difference in returns of the actual stocks, and also in the returns of the actuals minus ghosts. With equal-weight portfolios, the returns of the actuals minus ghosts in the *lowest* bin FF3 R^2 sorted bin is 1.183% (t-stat 3.796) per month, and is 0.533% (t-stat 4.299) in the *highest* R^2 bin, and we do find some statistical evidence that the *difference* of -0.651% (t-stat -2.123) is significant. However with value-weight portfolios, this statistical significance vanishes: the returns of the actuals minus ghosts in the *lowest* bin FF3 R^2 sorted bin is 0.588% (t-stat 2.788) per month, and is 0.389% (t-stat 3.212) in the *highest* R^2 bin, but the *difference* -0.199% (t-stat -0.882) is not statistically significant. The results are the same with FF5 R^2 sorted portfolios, where there is some statistical evidence to support that there is a difference in returns for the actuals minus ghosts between the lowest and highest R^2 bins in equal-weighted portfolios, but this difference vanishes with value-weighted portfolios. Thus unlike the elastic-net approach, we do *not* find a robust asset insurance premium in the cross-section of FF3- or FF5-based asset synchronicities.

[Table 15 about here.]

[Table 16 about here.]

The results of Tables 15 and 16 do not control for factor risks. We revisit the procedure of Section 3.2 and investigate whether the Fama-French z factor model alphas emerge out of portfolios that are sorted by FFz R^2's. Table 17 (equal-weight) and Table 18 (value-weight) show the results of the Fama-French 5 factor model regression estimates on post-formation FF5 R^2 sorted portfolios. [8] Recall the elastic-net counterparts are Table 7 (equal-weight) and Table 8 (value-weight), and there there is statistically significant alpha in those regressions. In equal-weighted portfolios of the actuals minus ghosts, we see that the Fama-French 5 alpha in the *lowest* FF5 R^2 bin is 0.875% per month (t-stat 3.778), the *highest* bin is 0.640% (t-stat 4.908), but their *difference* is statistically insignificant at -0.235% (t-stat -0.986). Likewise in value-weighted portfolios of the actuals minus ghosts, the Fama-French 5 alpha in the *lowest* FF5 R^2 bin is 0.448% (t-stat 2.397), the *highest* bin is 0.503% (t-stat 3.785), yet again their *difference* is statistically insignificant at 0.055% (t-stat 0.271). This "non-result" is particularly surprising because we use the Fama-French 5 factor alphas to evaluate precisely those portfolios that are sorted by FF5 R^2's. Indeed, one might have anticipated that this is particularly the type of setup that strongly favors the FF5 R^2 portfolio sorts.

[Table 17 about here.]

[Table 18 about here.]

5.2 Corroborating evidence of an unconditional average asset insurance premium

Overall, we fail to find an asset insurance premia in the cross-section of FFz constructed asset synchronicity. However, the *unconditional average* asset insurance premium *does* exist across all of elastic-net, FF3 and

[8]The results of the Fama-French 3 factor model based on post-formation FF3 and FF5 R^2 sorted portfolios are similar to the discussions here. We omit these results for brevity, and they are available upon request.

FF5 R^2 sorted portfolios. The "Avg" portfolio is the simple average return across all the 10 R^2 portfolio returns. Table 3 again shows the main elastic-net R^2 sort results of this paper, and there the equal-weighted average portfolio return of the actuals minus ghosts is 0.793% (t-stat 3.942) per month, and it is 0.575% (t-stat 4.333) for value-weights. The results for FF3 and FF5 R^2 sorted portfolios are remarkably similar. In Table 15 where we sort by FF3 R^2's, the equal-weighted portfolio return of the actual minus ghosts is 0.792% (t-stat 3.928) per month, and it is 0.557% (t-stat 3.895) with value-weights. In Table 16 where we sort by FF5 R^2's, the equal-weighted portfolio return of the actual minus ghosts is 0.791% (t-stat 3.927) per month, and it is 0.582% (t-stat 4.043) with value-weights. There should be no surprise that the results of FF3 and FF5 R^2 are nearly identical. However, it is surprising that these FFz results are nearly identical to our elastic-net R^2 sorted portfolios. After all, the construction of the ghosts (8) with the elastic-net approach compared to the construction of the ghosts (16) with the FFz approach are fundamentally different. This is evidence that the overall asset insurance premium effect does exist "on average" in this economy, and that this premium is not driven by any data-mining nature of our machine learning method.

We want to further emphasize that there *is* a statistical difference in the cross-section between the elastic-net and the FFz methods, even though there *is no* difference in the average. We take the two of the "Avg" portfolio returns of the actuals minus ghosts from the elastic-net, FF3 and FF5 methods and subtract them against each other. Table 19 shows the results. There is no statistical difference between the "average" portfolio returns of the actual minus ghosts between all three of the methods even after we control for the common factor risks. This result corroborates the pattern that we see in Tables 3, 15 and 16. Next, we take two of the "Hi - Lo" portfolio returns of the actual minus ghosts from the elastic-net, FF3 and FF5 methods and subtract them against each other. Table 20 shows the results. Even adjusting for the common factor risks, there remains a substantial difference in the portfolio composition between in the "Hi - Lo" bin between elastic-net and the FFz R^2 sorting methods. These two results further emphasize that an "average" asset insurance premium exists in the economy, regardless of whether we use an unstructured estimation method like the elastic-net, or a structured method like the FF3 and FF5 factor models to construct the ghost portfolios.

[Table 19 about here.]

[Table 20 about here.]

Following the discussions of Section 4, we also plot the cross-sectional average FFz R^2's in Figure 4. While the elastic-net R^2's tend to be higher than the FFz R^2's, their overall pattern appears to highly co-move with each other. This is again another piece of evidence that the elastic-net approach is detecting a similar average result as the workhorse FFz models.

5.3 Instability of portfolio weights in FF3 and FF5 ghosts

Although we fail to find an asset insurance premium in the cross-section of FFz R^2 portfolio sorts, it is nonetheless instructive to observe the behavior of the FFz ghost portfolios. Table 21 shows that stocks with higher FFz R^2 are associated with higher allocations the risky z factors in constructing its ghost portfolio. This qualitative pattern of higher R^2 and higher allocation into the risky assets in constructing the ghost portfolios is also observed in the elastic-net method. But this is effectively where the similarities between elastic-net and FFz OLS estimated coefficients end.

Quantitatively, the FFz estimated coefficients are widely dispersed. For instance, in the lowest FF5 R^2 bin, the average proportion of wealth allocated to the five risky factors in constructing the ghost portfolio of a given actual stock is 53.1%, but the 25-th percentile is -64.3% and the 75-th percentile is 161.6%, implying an interquartile range of $161.6\% - (-64.3\%) = 225.9\%$. Even in the highest FF5 R^2 bin, the average proportion is 145.6%, and the associated interquartile range remains wide at $285.3\% - 32.6\% = 252.7\%$. This means that within a given portfolio bin k as sorted by FF5 R^2, the ghost portfolio of a one stock i could involve a large short position in the risk free asset, while another stock j could involve a large short position in the five risky factors. This implies the risk and reward profiles of ghost portfolios within each FF5 R^2 bin could be highly heterogeneous, even though they are grouped by the same R^2 characteristic.

Figure 5 also plots the average equity proportion of the FFz ghosts against that of the elastic-net ghosts. The proportion of equity in the FFz ghosts fluctuate substantially more and are a higher magnitude than their elastic-net counterpart. This qualitatively confirms our above discussions of Table 21.

The ghost portfolios of the elastic-net are substantially more stable than that of the ghost portfolio constructed out of the FFz methods, as mentioned in Section 4.1. In all, that means the average and cross-sectional asset insurance premium constructed out of the elastic-net method are likely less likely to be eroded by trading frictions.

[Table 21 about here.]

6 Robustness check I: Asset insurance premium is not the equity risk premium

A potential concern for our results is that the ghosts in the lowest elastic-net R^2 bin are effectively only the risk free asset. Table 1 shows that low elastic-net R^2 is associated with matching low numbers of other risky assets to a given actual stock i. Observing the lowest elastic-net R^2 bin from Table 1(a), the average number of risky stocks in a ghost portfolio is 0, with an interquartile range of $1 - 0 = 1$. From the construction of the ghost portfolio of actual stock i in (8), it means many ghosts in this lowest possible bin are effectively just the risk free asset. This happens precisely when an actual stock i at month $t-1$ has an elastic-net estimated coefficient of $\hat{\beta}_{i,t-1} \equiv \mathbf{0}$. Several of our main results in Section 3 involve evaluating the statistical significance between the "Hi" R^2" and "Lo" R^2 bin. And if most of the ghosts in the "Lo" R^2 bin is just the risk free asset, then one may question whether the main results of Table 3 are just capturing the well-established equity premium of risky assets.

Table 22 shows the results where we subset only for actual stocks i whose ghost portfolios contain *at least one other risky asset*. In this construction, none of the ghost portfolios in any of the elastic-net R^2 bins can be the risk free asset. With equal-weight portfolios, we see that the asset insurance premium is 1.348% (t-stat 4.664) per month for the *least* synchronized stocks, almost monotonically decreases to 0.370% (t-stat 3.401) for the *most* synchronized stocks, and the *difference* is -0.978% (t-stat -3.422). With value-weight portfolios, the asset insurance premium is 0.906% (t-stat 4.875) for the *least* synchronized stocks, almost monotonically decreases to 0.335% (t-stat 2.999) for the *most* synchronized stocks, and the *difference* is -0.571% (t-stat -3.061). While the numerical magnitudes are slightly smaller in this subsetted case, the overall existence of the asset insurance premium is as economically prominent as in the full result of Table 3.

Tables 23 and 24 are the counterparts to Tables 7 and 8, respectively. We see after controlling for the FF5 risk factors and subsetting for minimum number of risky assets in the ghost portfolios of each stock, the cross-sectional asset insurance premium, and the unconditional average asset insurance premium still strongly exist.

[Table 22 about here.]

[Table 23 about here.]

[Table 24 about here.]

7 Robustness check II: Controlling for known risk factors

Section B.2 of the Appendix reports bivariate dependent portfolio sort results. We show that after controlling for risk factors (i.e. total and idiosyncratic volatility, FF3 and FF5 R^2's, size, book-to-market, dollar volume liquidity, Amihud's illiquidity, short-term reversal and momentum) we still find strong evidence that stocks with (low) high asset synchronicity earn a high (low) asset insurance premium in the cross-section.

8 Conclusion

We theoretically prove and empirically show the existence of an *asset insurance premium*, which represents the residual factor risks of a stock. The three core findings of the paper are: (i) There is an unconditional *average* asset insurance premium of 0.575% per month; (ii) There is an asset insurance premium in the *cross-section* of asset synchronicity, where lowly (highly) synchronized stocks earn a 0.976% (0.305%) per month; and (iii) asset synchronicity is highly countercyclical, and negatively related with shocks to consumption. We emphasize that asset insurance premium is distinct from the well-known equity risky premium.

Our results are robust to controlling for the well-known Fama-French three and five factors, and robust to bivariate dependent portfolio sorts against value, size, total and idiosyncratic volatility, dollar volume liquidity and Amihud's illiquidity measure. The robustness of a cross-sectional asset insurance premium across asset synchronicity is particularly crucial because it shows our results strongly differentiate against known results of idiosyncratic volatility as a priced factor. Compared to traditional linear based regression methods, we show in this paper the efficacy of machine learning methods, such as the elastic-net estimator, to illuminate new cross-sectional empirical asset pricing facts. While we have strongly advocated using the elastic-net to project a risky asset's returns onto the returns span of other assets, a potential avenue for future research could be to design better projection methods, specialized for financial economic data, that are inspired by the machine learning literature.

References

Adrian, T. and M. K. Brunnermeier (2016): "CoVaR," *American Economic Review*, 106, 1705–1741.

Amihud, Y. (2002): "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, 5, 31–56.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61, 259–299.

Bai, J. and S. Ng (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131, 507–537.

Beber, A., M. W. Brandt, and K. A. Kavajecz (2008): "Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market," *The Review of Financial Studies*, 22, 925–957.

Black, F., M. C. Jensen, and M. Scholes (1972): "The capital asset pricing model: Some empirical tests," *Studies in the theory of capital markets*, 81, 79–121.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989): "Empirical tests of the consumption-orientated CAPM," *Journal of Finance*, 44, 231–262.

Campbell, J. Y. (2003): "Consumption-based asset pricing," *Handbook of the Economics of Finance*, 1, 803–887.

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

Chinco, A., A. D. Clark-Joseph, and M. Ye (2018): "Sparse Signals in the Cross-Section of Returns," *Journal of Finance*, 0.

Cochrane, J. H. (2005): *Asset Pricing: Revised Edition*, Princeton University Press.

Fama, E. F. and K. R. French (1992): "The cross-section of expected stock returns," *Journal of Finance*, 47, 427–465.

——— (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3 – 56.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1 – 22.

Feng, G., G. Stefano, and D. Xiu (2017): "Taming the Factor Zoo," *Working Paper*.

Freyberger, J., A. Neuhierl, and M. Weber (2017): "Dissecting Characteristics Nonparametrically," Working Paper 23227, National Bureau of Economic Research.

Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst (2006): "Pairs trading: Performance of a relative-value arbitrage rule," *The Review of Financial Studies*, 19, 797–827.

Gu, S., B. Kelly, and D. Xiu (2018): "Empirical Asset Pricing via Machine Learning," *Working Paper*.

Hansen, L. P. and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–853.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "... and the Cross-Section of Expected Returns," *Review of Financial Studies*, 29, 5–68.

Hedegaard, E. and R. J. Hodrick (2016): "Estimating the risk-return trade-off with overlapping data inference," *Journal of Banking and Finance*, 67, 135 – 145.

James, M. (1978): "The Generalised Inverse," *The Mathematical Gazette*, 62, 109–114.

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

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

Jin, L. and S. C. Myers (2006): "R2 around the world: New theory and new tests," *Journal of Financial Economics*, 79, 257 – 292.

Krauss, C. (2017): "Statistical arbitrage pairs trading strategies: Review and outlook," *Journal of Economic Surveys*, 31, 513–545.

Lamont, O. A. (2001): "Economic tracking portfolios," *Journal of Econometrics*, 105, 161–184.

Lehmann, B. N. (1990): "Fads, martingales, and market efficiency," *Quarterly Journal of Economics*, 105, 1–28.

Lintner, J. (1965): "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *The Review of Economics and Statistics*, 47, 13–37.

Longstaff, F. A. (2004): "The Flight-to-Liquidity Premium in US Treasury Bond Prices," *The Journal of Business*, 77, 511–526.

Morck, R., B. Yeung, and W. Yu (2000): "The information content of stock markets: why do emerging markets have synchronous stock price movements?" *Journal of Financial Economics*, 58, 215 – 260, special Issue on International Corporate Governance.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 703–708.

Pástor, L. and R. F. Stambaugh (2003): "Liquidity risk and expected stock returns," *Journal of Political Economy*, 111, 642–685.

Penrose, R. (1955): "A generalized inverse for matrices," in *Mathematical proceedings of the Cambridge philosophical society*, Cambridge University Press, vol. 51, 406–413.

Ross, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Sharpe, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

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

Wurgler, J. and E. Zhuravskaya (2002): "Does arbitrage flatten demand curves for stocks?" *The Journal of Business*, 75, 583–608.

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

A Appendix

A.1 Discussions and proofs to Section 1

Throughout this section, we use $\boldsymbol{I}_n$ to denote the $n \times n$ identity matrix, $\mathbf{0}_n$ to denote the zero vector in $\mathbb{R}^n$, and $\mathbf{O}_{n\times m}$ to denote the $n \times m$ dimensional matrix of all zeros. All quantities are real-valued in this section.

In what follows, we will need the definition of a *pseuoinverse* of a matrix.

Definition A.1. If A is an $m \times n$ real-valued matrix, then the *pseudoinverse* of A is defined as an $n \times m$ real-valued matrix A^+ that satisfies all of the four following *Moore-Penrose* conditions:

1. $AA^+A = A$;
2. $A^+AA^+ = A^+$;
3. $(AA^+)^\top = AA^+$; and
4. $(A^+A)^\top = A^+A$

Moreover, it can be shown that the pseuoinverse A^+ for a matrix A always exists and is unique.

We will also need the *image space* and *kernel space* of an $m \times n$ real-valued matrix A,

$$\mathrm{Im}(A) := \{Ax \in \mathbb{R}^m : x \in \mathbb{R}^n\}, \tag{18}$$

$$\mathrm{Ker}(A) := \{x \in \mathbb{R}^n : Ax = \mathbf{0}_m\}. \tag{19}$$

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In what follows, we will make our arguments "ω-by-ω", meaning all statements are to be read as $\mathbb{P}$-almost surely wherever a random variable is involved. Consider a single period economy. Suppose there are $N+1$ risky assets, a single risk free asset, and the excess returns of all risky assets are governed by a K set of factors. Ross (1976) argues for the equilibrium existence of such factors. That is, suppose the excess returns of asset $i = 1, \ldots, N+1$ is given by (1a), where $\boldsymbol{\beta}_i \in \mathbb{R}^K$ is asset i's factor loadings onto the factors, $\boldsymbol{F} \in \mathbb{R}^K$ is the vector of random factors, and $\varepsilon_i \in \mathbb{R}$ is the random idiosyncratic returns of asset i.

Define the stacked vectors,

$$\mathbf{a}_i := \begin{bmatrix} \alpha_i \\ \boldsymbol{\beta}_i \end{bmatrix}, \quad \boldsymbol{f} := \begin{bmatrix} 1 \\ \boldsymbol{F} \end{bmatrix},$$

so that (1a) can be equivalently expressed as,

$$R_i = \mathbf{a}_i^\top \boldsymbol{f} + \varepsilon_i \tag{20}$$

Let's stack (20) together for all of other risky assets other than asset i. That is, consider

$$\boldsymbol{R}_{-i} = \boldsymbol{A}_{-i}\boldsymbol{f} + \boldsymbol{\varepsilon}_{-i}, \tag{21}$$

where

$$\boldsymbol{R}_{-i} := \begin{bmatrix} R_1 \\ \vdots \\ R_{i-1} \\ R_{i+1} \\ \vdots \\ R_{N+1} \end{bmatrix}, \quad \boldsymbol{A}_{-i} := \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_{i-1}^\top \\ \mathbf{a}_{i+1}^\top \\ \vdots \\ \mathbf{a}_{N+1}^\top \end{bmatrix} = \begin{bmatrix} \alpha_1 & \boldsymbol{\beta}_1^\top \\ \vdots & \vdots \\ \alpha_{i-1} & \boldsymbol{\beta}_{i-1}^\top \\ \alpha_{i+1} & \boldsymbol{\beta}_{i+1}^\top \\ \vdots & \vdots \\ \alpha_{N+1} & \boldsymbol{\beta}_{N+1}^\top \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\alpha}_{-i} & \boldsymbol{B}_{-i} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_{-i} := \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{i-1} \\ \varepsilon_{i+1} \\ \vdots \\ \varepsilon_{N+1} \end{bmatrix},$$

so $\boldsymbol{R}_{-i} \in \mathbb{R}^N$ is the vector of excess returns of all other risky assets except asset i, $\boldsymbol{A}_{-i} \in \mathbb{R}^{N\times(K+1)}$ is the matrix of factor loadings that is block partitioned into the $\boldsymbol{\alpha}_{-i} \in \mathbb{R}^N$ vector and the $\boldsymbol{B}_{-i} \in \mathbb{R}^{N\times K}$ matrix, and $\boldsymbol{\varepsilon}_{-i} \in \mathbb{R}^N$ is the vector of idiosyncratic returns. To avoid trivialities, we will assume throughout that $\boldsymbol{B}_{-i} \neq \mathbf{O}_{N\times K}$, although we will consider the case where $\boldsymbol{\alpha}_{-i}$ is zero or non-zero.

We will also require the following technical assumption in the proof of Theorem 1.1.

Assumption A.1 (Subspace inclusion condition)**.** *Assume for all assets i, we have that*

$$\boldsymbol{A}_{-i}\boldsymbol{A}_{-i}^{+}(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}) = \boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}, \quad \mathbb{P}\text{-a.s.} \tag{22}$$

It can be shown that this condition is equivalent to requiring that $\operatorname{Im}(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i})$ *is a subset of* $\operatorname{Im}(\boldsymbol{A}_{-i})$.

This assumption says that the return realizations of risky asset less their idiosyncratic return realizations $\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}$ can be expressed as a linear combination of the factor loadings $\boldsymbol{A}_{-i}$ of the economy, and the return realizations of some factors. In other words, this is the technical condition that there is no model misspecification in (1a) (i.e. the only possible factors in the economy are $\boldsymbol{f}$ and nothing else).

Proof of Theorem 1.1. The idea is to view (21) as a system of linear equations and "solve" for the factors $\boldsymbol{f}$. But $N > K+1$ meaning there are more assets in the economy than factors, this means the matrix $\boldsymbol{A}_{-i}$ is not square, and so an inverse to $\boldsymbol{A}_{-i}$ does not exist. However, by using the *pseudoinverse*, we can nonetheless still solve (21) for $\boldsymbol{f}$.

Rearranging (21) as $\boldsymbol{A}_{-i}\boldsymbol{f} = \boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}$ so that it has the form of a system of (overdetermined) linear equations. Provided Assumption A.1 holds, a solution exists for this system of linear equations. [9] In particular, we can use the pseudoinverse to solve for $\boldsymbol{f}$,

$$\boldsymbol{f} = \boldsymbol{A}_{-i}^{+}(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}) + [\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}]\boldsymbol{w}_{-i}, \tag{23}$$

where $\boldsymbol{w}_{-i} \in \mathbb{R}^{K+1}$ is an arbitrary random vector. [10] A standard result of pseudoinverse has that $\operatorname{Ker}(\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}) = \operatorname{Im}(\boldsymbol{A}_{-i}^{+})$. To avoid technical issues in the proof of Corollary 1.3, we will specifically pick a vector $\boldsymbol{w}_{-i} \notin \operatorname{Im}(\boldsymbol{A}_{-i}^{+})$ so that $[\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}]\boldsymbol{w}_{-i} \neq \mathbf{0}_K$, $\mathbb{P}$-a.s.

Substituting (23) into (1a), we have that,

$$\begin{aligned}
R_i &= \mathbf{a}_i^{\top}\boldsymbol{f} + \varepsilon_i \\
&= \mathbf{a}_i^{\top}\left(\boldsymbol{A}_{-i}^{+}(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i}) + [\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}]\boldsymbol{w}_{-i}\right) + \varepsilon_i \\
&= \mathbf{a}_i^{\top}\boldsymbol{A}_{-i}^{+}\boldsymbol{R}_{-i} - \mathbf{a}_i^{\top}\boldsymbol{A}_{-i}^{+}\boldsymbol{\varepsilon}_{-i} + \mathbf{a}_i^{\top}[\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}]\boldsymbol{w}_{-i} + \varepsilon_i
\end{aligned}$$

We identify the terms $\mathbf{b}_i := (\mathbf{a}_i^{\top}\boldsymbol{A}_{-i}^{+})^{\top}$ and $\boldsymbol{\Phi}_i := [\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}]\boldsymbol{w}_{-i}$ to get the decomposition (1b). $\square$

Remark A.2. In the proof of Theorem 1.1, we have simply relied on Assumption A.1. It is worth noting that we explicitly do *not* assume any rank condition on the matrix $\boldsymbol{A}_{-i}$. Since $\boldsymbol{A}_{-i}$ is a $N \times (K+1)$

[9] Penrose (1955) shows that this condition is actually sufficient and necessary for the existence of a solution to the overdetermined system of linear equations. See also James (1978) for a succinct introduction to solving a system of linear equations with pseudoinverses.

[10] The randomness of $\boldsymbol{w}_{-i}$ can be more formally viewed as follows. Since (21) is an equation of random variables in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then (21) must hold also for each $\omega \in \Omega$,

$$\boldsymbol{R}_{-i}(\omega) = \boldsymbol{A}_{-i}\boldsymbol{f}(\omega) + \boldsymbol{\varepsilon}_{-i}(\omega).$$

Then proceeding again via the pseuoinverse as in (23), we must then also have

$$\boldsymbol{f}(\omega) = \boldsymbol{A}_{-i}^{+}(\boldsymbol{R}_{-i} - \boldsymbol{\varepsilon}_{-i})(\omega) + [\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}]\boldsymbol{w}_{-i}(\omega).$$

Since the above holds for all $\omega \in \Omega$, so $\boldsymbol{w}_{-i}$ must be a $\mathcal{F}$-measurable function, and thus is a random variable.

matrix, the rank-nullity theorem suggests that $\text{rank}(\boldsymbol{A}_{-i}) \leq \min\{N, K+1\} = K+1$. Thus there are two important cases to consider for the rank of $\text{rank}(\boldsymbol{A}_{-i})$: the case of when its columns are linearly independent, so $\text{rank}(\boldsymbol{A}_{-i}) = K+1$; and the case when its columns are linearly dependent, so $\text{rank}(\boldsymbol{A}_{-i}) < K+1$. If $\text{rank}(\boldsymbol{A}_{-i}) = K+1$, then by properties of the pseudoinverse, it can be shown that $\boldsymbol{A}_{-i} = (\boldsymbol{A}_{-i}^{\top}\boldsymbol{A}_{-i})^{-1}\boldsymbol{A}_{-i}^{T}$, and so in this case we get that $\boldsymbol{\Phi}_i = \boldsymbol{0}_{K+1}$. At this point in the discussion, Theorem 1.1 does permit $\boldsymbol{\Phi}_i$ to (uninterestingly) equal to zero. But Corollary 1.3 provides a mild economic condition for which $\boldsymbol{\Phi}_i$ will be non-zero.

A.2 Discussions and details of the elastic-net estimator

We introduce the optimization problem of the *elastic-net* estimator developed by Zou and Hastie (2005). Let y be a $T \times 1$ vector, $\boldsymbol{X}$ be a $T \times N$ matrix and let β be a $N \times 1$ vector. Consider the optimization problem,

$$\hat{\beta}_{\text{EN}} = \underset{\beta \in \mathbb{R}^N}{\arg\min} \left\{ \frac{1}{2T} \|y - \boldsymbol{X}\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right\}, \quad \lambda_1, \lambda_2 \geq 0, \tag{24}$$

where $\|x\|_1 := \sum_{j=1}^N |x_j|$ is the L^1-norm on $\mathbb{R}^N$, and $\|x\|_2 := \sqrt{\sum_{j=1}^N x_j^2}$ is the L^2-norm on $\mathbb{R}^N$. In our application, y will be a time series vector of T days of returns of a particular stock, and $\boldsymbol{X}$ will be the concatenation of the time series of N number of other stocks. Note this means there are $N+1$ total number of stocks.

The solution $\hat{\beta}_{\text{EN}}$ is called the *elastic-net* estimator. This estimator encompasses the special cases of the *ordinary least squares (OLS)* estimator (when $\lambda_1 = 0, \lambda_2 = 0$), *least absolute shrinkage and selection operator (LASSO)* estimator of Tibshirani (1996) (when $\lambda_1 > 0, \lambda_2 = 0$), and the *ridge* estimator (when $\lambda_1 = 0, \lambda_2 > 0$). The hyperparameters λ_1, λ_2 control the strength of the L^1- and L^2-norm penalties, respectively. In this paper when we refer to the elastic-net estimator, we always refer to the case when λ_1, λ_2 are both strictly positive. In our actual implementation, we use a 3-fold *cross-validation* procedure to empirically select the hyperparameters $\lambda_1, \lambda_2 > 0$. [11]

In Theorem 1.1, we motivated the need to *linearly* regress a stock return R_i onto the returns $\boldsymbol{R}_{-i}$ of all other stocks. The elastic-net estimator can actually be seen as a linear regression problem with constraints. By Lagrange-duality, (24) is identical to the following constrained least squares problem,

$$\begin{aligned} \min_{\beta \in \mathbb{R}^N} \quad & \frac{1}{2T} \|y - \boldsymbol{X}\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \leq \eta_1, \\ & \|\beta\|_2^2 \leq \eta_2, \end{aligned} \tag{25}$$

for some $\eta_1, \eta_2 \geq 0$ that is dependent on the values of $\lambda_1, \lambda_2 \geq 0$. Most references in the statistical literature (e.g. Zou and Hastie (2005)) prefers the unconstrained Lagrangian-form (24) over the constrained optimization form (25) for theoretical and computational reasons. Here, we explicitly draw out the "linear nature" of the elastic-net estimator via (25) because Theorem 1.1 indeed predicts a linear relationship.

[11] To reduce the already lengthy computational time in estimating (24), we cross-validate for only one hyperparameter rather than two by making the following simplifying assumption on λ_1, λ_2. We set $\lambda_1 = \lambda\ell$ and $\lambda_2 = \frac{1}{2}\lambda(1-\ell)$, and set $\ell = 1/2$, and only cross-validate for the single $\lambda > 0$ parameter.

A.3 Additional discussions on the estimation procedure

A.3.1 Intercept estimation, stale prices and sparsity

We deliberately do *not* estimate an intercept in (24). [12] This is to avoid attributing a price with very stale prices with high R^2.

Imagine a given stock i has a vector of twelve months' worth daily returns $y_{i,t-1}$, where all the daily returns are almost all 0's except on a handful of days. As an extreme, suppose all other stocks has a return matrix $\boldsymbol{X}_{i,t-1}$ of (5) with rank $D_{i,t-1}$. Note that we do not impose nor check for any such conditions in our implementation. Then using the elastic-net estimator without intercept (24), the squared error term (i.e. numerator of (6)) would be high, and thus leading to an overall low $\mathsf{R}^2{}_{i,t-1}$. Observe that in this case $\hat{\beta}_{i,t-1} = \mathbf{0}$ is not necessarily an optimal solution because $\boldsymbol{X}_{i,t-1}$ is of rank $D_{i,t-1}$ while β is $N_{t-1} \times 1$, and we have that $D_{i,t-1} \ll N_{t-1}$. This means there could exist some sparse $\hat{\beta}_{i,t-1} \neq \mathbf{0}$ that achieves a smaller value in (24) than that of $\beta = \mathbf{0}$, namely when $||y_{i,t-1} - \boldsymbol{X}_{i,t-1}\hat{\beta}_{i,t-1}||_2^2 + \lambda_1||\hat{\beta}_{i,t-1}||_1 + \lambda_2||\hat{\beta}_{i,t-1}||_2^2 \ll ||y_{i,t-1}||_2^2$, especially when the penalty weights λ_1, λ_2 are small. In contrast, if one were to use the elastic-net estimator *with* an intercept, then setting $\hat{\beta}_{i,t-1} = \mathbf{0}$ with intercept $\hat{c} \approx \mathbf{0}$ *is* an optimal solution. As a result, this would lead to the numerator in (6) to approximately equal to zero, and thus resulting in a high R^2. We want to *avoid* the latter case in our results.

In this paper, we want to exclusively reserve "high R^2" to mean stock i can be well explained by other risky assets, and "low R^2" to mean stock i cannot be explained by other risky assets.

A.3.2 Technical details of the importance of parsimony

A general form of a machine learning estimator has the form $y = g(\mathbf{X}; \theta, \lambda) + \varepsilon$, where y is the response variable, g could be a parameterized or non-parametric function, $\mathbf{X}$ is the set of regressors, θ is parameterizes g, λ is a hyperparameter, and ε is a nuisance parameter. Generally speaking, the relationship between $\mathbf{X}$, θ and λ could be highly non-linear. The data is typically split into three sets $[y_{\text{train}}, y_{\text{validate}}, y_{\text{test}}]$ and $[\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{validate}}, \mathbf{X}_{\text{test}}]$. For a given hyperparameter λ, the machine learning method uses the *training set* $(y_{\text{train}}, \mathbf{X}_{\text{train}})$ to fit the data to find an optimal parameter $\hat{\theta}(\lambda)$. Using the *validation set* $(y_{\text{valid}}, \mathbf{X}_{\text{valid}})$, the method then finds an optimal $\hat{\lambda}$ that fits the validation data, and a trained model is then given by the parameters $\hat{\theta}(\hat{\lambda})$ and $\hat{\lambda}$. Finally, the forecast accuracy of the model is tested against the *testing set* $(y_{\text{test}}, \mathbf{X}_{\text{test}})$ by comparing against the predicted value $\hat{y} = g(\mathbf{X}_{\text{train}}; \hat{\theta}(\hat{\lambda}), \hat{\lambda})$ against its realization y_{train}.

In our context, if we were to apply a general machine learning method to asset i, then we would still use the same response variable and regressors as in (5). However, the fitted value would be of very little use to us when constructing portfolios from a finance perspective. There are two issues. Firstly, we are *not* using a forecast value $\hat{y}$ to construct portfolios. We are using the fitted coefficients $\hat{\theta}(\lambda) = \hat{\beta}_{\text{EN},i,t-1}$ as investment weights for a portfolio. In order to use fitted coefficients as investment weights, then it is almost *necessary* that these fitted coefficients have a clear and linear relationship to the one-month ahead returns $\boldsymbol{X}_{\text{train}} = \boldsymbol{R}_{-i,t}$. The OLS, LASSO and elastic-net certainly have this property, as seen in (8). However, a general machine learning method does *not* have this linear relationship between the fitted coefficients and its out-of-sample regressors. This non-linearity between fitted coefficients and its out-of-sample regressors explicitly prevent us from using these general machine learning methods in constructing a portfolio.

---

[12] The elastic-net estimator that contains an intercept is given by,

$$\hat{c}_{\text{EN}}, \hat{\beta}_{\text{EN}} \in \underset{c \in \mathbb{R}, \beta \in \mathbb{R}^p}{\arg\min} \left\{ ||y - c - \boldsymbol{X}\beta||_2^2 + \lambda_1 \, ||\beta||_1 + \lambda_2 \, ||\beta||_2^2 \right\}, \quad \lambda_1, \lambda_2 \geq 0,$$

where by convention, L^1- and L^2-penalties are only applied on the β coefficients and not on the intercept c.

B Internet Appendix

B.1 Extended proofs to Section 1

Here we make the term *correlation* precise, since we are dealing with random quantities ω-by-ω on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and not in expectation.

Definition B.1 (Relative factor correlation)**.** We say there are L number of *uncorrelated* factors (relative to asset i) if the factor loadings matrix of other assets $\boldsymbol{B}_{-i}$ has rank L. We also say that there are $K - L \geq 0$ number of *correlated factors.* [13]

Recall the matrix $\boldsymbol{B}_{-i}$ has dimensions $N \times K$ and rank of at most $\min\{N, K\} = K$. If that matrix actually has rank $L < K$, then economically it means the N assets can be uniquely determined by just L number of factors. In other words, the other $K - L$ number of factors are in effect redundant. Econometrically, we can view this redundancy akin to these $K - L$ factors being dependent, as opposed to being independent.

Assumption B.1 (No redundant assets)**.** *Given any asset i that is not a factor itself, assume that*

(i) $\boldsymbol{\beta}_i \notin \mathrm{Im}(\boldsymbol{B}_{-i}^\top)$.

If $[\alpha_1, \ldots, \alpha_i, \ldots, \alpha_{N+1}]^\top \neq \mathbf{0}_{N+1}$*, further assume that*

(ii) $\boldsymbol{\alpha}_{-i} \notin \mathrm{Ker}(\boldsymbol{B}_{-i}^\top)$*; and*

(iii) $||\boldsymbol{\beta}_i||_2 / |\alpha_i| > ||\boldsymbol{B}_{-i}^\top||_2 / ||\boldsymbol{\alpha}_{-i}||_2$

Here $||\cdot||_2$ *denotes both the Euclidean norm for vectors, and also the associated operator norm on the space of real matrices.* [14]

Condition (i) is the most important economic assumption. Indeed, when markets are efficient the only condition that we require is (i). Conditions (ii) and (iii) are technical which are applicable only if there's some form of market inefficiency, where $\alpha_i \neq 0$ and $\alpha_j \neq 0$ for some $j \neq i$. Condition (i) simply requires that the factor loading of asset i cannot equal to some linear combination of the factor loadings of other assets. Economically, this requires that asset i cannot be a redundant asset. Note that equivalently, if $\boldsymbol{\beta}_i \notin \mathrm{Im}(\boldsymbol{B}_{-i}^\top)$, then we have that

$$\boldsymbol{\beta}_i \notin \mathrm{span}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{i-1}, \boldsymbol{\beta}_{i+1}, \ldots, \boldsymbol{\beta}_{N+1}\}$$

Condition (ii) is a technical condition that says the beta loadings multiplied by the alphas cannot cancel each other out. Condition (iii) is an intuitive condition when asset i has a non-zero intercept. It requires that the relative size of asset i's beta loadings to alpha to be strictly larger than the relative size of all other assets' beta loadings to alpha.

[13] If two mean-zero random vectors $X, Y \in \mathbb{R}^K$ are uncorrelated (in expectations) then $\mathbb{E}[\langle X, Y\rangle] = 0$. This can be seen as an orthogonality condition in L^1. The definition we have taken here is a strong necessary condition of orthogonality. To see this, if the set of vectors $\{X(\omega), Y(\omega)\}$ is orthonormal in $\mathbb{R}^K$ for all ω, then $\langle X, Y\rangle = 0$ $\mathbb{P}$-a.s., so necessarily we have that $\mathbb{E}[\langle X, Y\rangle] = 0$. Non-trivial orthonormal sets are necessarily linearly independent.

[14] Specifically, if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then $||x||_2 := \sqrt{x_1^2 + \ldots + x_n^2}$. And the operator norm for an $m \times n$ real matrix is defined as,

$$||A||_2 := \sup\left\{\frac{||Ax||_2}{||x||_2} : x \in \mathbb{R}^n, x \neq \mathbf{0}_n\right\}$$

Proof of Corollary 1.3. Fix any asset i. There are two cases to consider: asset i is a factor itself; or asset i is not a factor. Suppose asset i is one of the factors, so $R_i = F_k$ for some $k = 1, \dots, K$. So from (1a) we necessarily have $\alpha_i = 0$, $\boldsymbol{\beta}_i = \boldsymbol{e}_k$, and $\varepsilon_i = 0$. So in (1b), we can simply *define* $\mathbf{b}_i \equiv \mathbf{0}_N$ and $\boldsymbol{\Phi}_i^\top \equiv [1, \boldsymbol{F}]$. Given that the factors generate a positive excess return, then the asset insurance premium in this case is simply the factor return itself, which is positive.

The bulk of the proof is on the more involved case where asset i is not one of the factors. Observe that for all $\omega \in \Omega$,

$$\mathbf{a}_i^\top \boldsymbol{\Phi}_i(\omega) = \mathbf{a}_i^\top (\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i}) \boldsymbol{w}_{-i}(\omega) = \mathbf{a}_i^\top (\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i})(\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i}) \boldsymbol{w}_{-i}(\omega),$$

where the last equality follows from the projection properties of the pseudoinverse. Thus to show that $\mathbf{a}_i^\top \boldsymbol{\Phi}_i(\omega)$ is non-zero almost surely, then we need to first show that $\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i}$ is non-zero. Then we need to show that both $\mathbf{a}_i^\top (\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i})$ and $(\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i}) \boldsymbol{w}_{-i}(\omega)$ are non-zero. But by construction in the proof of Theorem 1.1, we have picked the random vector $\boldsymbol{w}_{-i} \notin \mathrm{Im}(\boldsymbol{A}_{-i}^+)$ so that $(\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i}) \boldsymbol{w}_{-i}(\omega) \neq \mathbf{0}_{K+1}$.

Observe that,

$$\begin{aligned}
\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i} &= \boldsymbol{I}_{K+1} - \begin{bmatrix} \boldsymbol{\alpha}_{-i}^+ \\ \boldsymbol{B}_{-i}^+ \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{-i} & \boldsymbol{B}_{-i} \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0}_K^\top \\ \mathbf{0}_K & \boldsymbol{I}_K \end{bmatrix} - \begin{bmatrix} \boldsymbol{\alpha}_{-i}^+ \boldsymbol{\alpha}_{-i} & \boldsymbol{\alpha}_{-i}^+ \boldsymbol{B}_{-i} \\ \boldsymbol{B}_{-i}^+ \boldsymbol{\alpha}_{-i} & \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} \end{bmatrix} \\
&= \begin{cases} \begin{bmatrix} 1 & \mathbf{0}_K^\top \\ \mathbf{0}_K & \boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} \end{bmatrix}, & \text{if } \boldsymbol{\alpha}_{-i} = \mathbf{0}_N \\ \begin{bmatrix} 0 & -\frac{\boldsymbol{\alpha}_{-i}^\top}{||\boldsymbol{\alpha}_{-i}||_2^2} \boldsymbol{B}_{-i} \\ -\boldsymbol{B}_{-i}^+ \boldsymbol{\alpha}_{-i} & \boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} \end{bmatrix}, & \text{if } \boldsymbol{\alpha}_{-i} \neq \mathbf{0}_N \end{cases}
\end{aligned} \tag{26}$$

where (26) follows from the fact that the pseudoinverse of a real-valued vector $\boldsymbol{x} \in \mathbb{R}^n$ is,

$$\boldsymbol{x}^+ = \begin{cases} \mathbf{0}_n^\top, & \text{if } \boldsymbol{x} = \mathbf{0}_n, \\ \frac{\boldsymbol{x}^\top}{||\boldsymbol{x}||_2^2}, & \text{if otherwise,} \end{cases}$$

and $||\boldsymbol{x}||_2 := \sqrt{x_1^2 + \ldots + x_n^2}$ is the Euclidean norm on $\mathbb{R}^n$. Thus there are two cases in (26) to consider, depending on whether there is no mispricing in the economy, or when there is mispricing in the economy.

<u>*Case (a)* $[\alpha_1, \dots, \alpha_i, \dots, \alpha_{N+1}]^\top = \mathbf{0}_{N+1}$</u>: Since $\alpha_i = 0$, then from the first case of (26)

$$\mathbf{a}_i^\top (\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^+ \boldsymbol{A}_{-i}) = \begin{bmatrix} 0 & \boldsymbol{\beta}_i^\top \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_K^\top \\ \mathbf{0}_K & \boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} \end{bmatrix} = \boldsymbol{\beta}_i^\top (\boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i}). \tag{27}$$

Thus if $\boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} = \mathbf{O}_{K \times K}$, then the asset insurance premium of asset i is necessarily zero. It can be shown when the $N \times K$ matrix $\boldsymbol{B}_{-i}$ has full column rank, so that $\mathrm{rank}(\boldsymbol{B}_{-i}) = K$, then it can be shown that its pseudoinverse simplifies to $\boldsymbol{B}_{-i}^+ = (\boldsymbol{B}_{-i} \boldsymbol{B}_{-i}^\top)^{-1} \boldsymbol{B}_{-i}$, and so $\boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} = \mathbf{O}_{K \times K}$. But by hypothesis and Definition B.1, $\mathrm{rank}(\boldsymbol{A}_{-i}) < K$ when some of the factors are correlated, so $\boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i} \neq \mathbf{O}_{K \times K}$. As a general property of pseudoinverses, it can be shown that (e.g. James (1978)) $\mathrm{Ker}(\boldsymbol{I}_K - \boldsymbol{B}_{-i}^\top (\boldsymbol{B}^\top)_{-i}^+) = \mathrm{Im}(\boldsymbol{B}_{-i}^\top)$. Since asset i is a non-redundant asset, then Assumption B.1(i) holds. This means $\boldsymbol{\beta}_i \notin \mathrm{Ker}(\boldsymbol{I}_K - \boldsymbol{B}_{-i}^\top (\boldsymbol{B}_{-i}^\top)^+)$, and hence $\boldsymbol{\beta}_i^\top (\boldsymbol{I}_K - \boldsymbol{B}_{-i}^+ \boldsymbol{B}_{-i}) \neq \mathbf{0}_K^\top$. Thus (27) is non-zero.

<u>*Case (b)* $[\alpha_1, \ldots, \alpha_i, \ldots, \alpha_{N+1}]^\top \neq \mathbf{0}_{N+1}$</u>: Let's consider the second case of (26). Observe that in this case,

$$\begin{aligned}
\mathbf{a}_i^\top (\boldsymbol{I}_{K+1} - \boldsymbol{A}_{-i}^{+}\boldsymbol{A}_{-i}) &= \begin{bmatrix}\alpha_i & \boldsymbol{\beta}_i^\top\end{bmatrix} \begin{bmatrix} 0 & -\frac{\boldsymbol{\alpha}_{-i}^\top}{||\boldsymbol{\alpha}_{-i}||_2^2}\boldsymbol{B}_{-i} \\ -\boldsymbol{B}_{-i}^{+}\boldsymbol{\alpha}_{-i} & \boldsymbol{I}_K - \boldsymbol{B}_{-i}^{+}\boldsymbol{B}_{-i}\end{bmatrix} \\
&= \begin{bmatrix}-\boldsymbol{\beta}_i^\top \boldsymbol{B}_{-i}^{+}\boldsymbol{\alpha}_{-i} & -\alpha_i \frac{\boldsymbol{\alpha}_{-i}^\top}{||\boldsymbol{\alpha}_{-i}||_2^2}\boldsymbol{B}_{-i} + \boldsymbol{\beta}_i^\top (\boldsymbol{I}_K - \boldsymbol{B}_{-i}^{+}\boldsymbol{B}_{-i})\end{bmatrix} \qquad (28)
\end{aligned}$$

Now we have two further subcases: whether $\alpha_i = 0$ or $\alpha_i \neq 0$. If $\alpha_i = 0$, then the second element of (28) remains non-zero by the arguments presented in Case (i). This is sufficient to conclude (28) non-zero.

The second subcase of $\alpha_i \neq 0$ is more involved. Again, it suffices to show that the second element of (28) is non-zero. To that end, we take the transpose for convenience and calculate,

$$\begin{aligned}
\left|\left|(\boldsymbol{I}_K - \boldsymbol{B}_{-i}^\top(\boldsymbol{B}_{-i}^\top)^{+})\boldsymbol{\beta}_i - \alpha_i \boldsymbol{B}_{-i}^\top \frac{\boldsymbol{\alpha}_{-i}}{||\boldsymbol{\alpha}_{-i}||_2^2}\right|\right|_2 &\geq \left|\left|(\boldsymbol{I}_K - \boldsymbol{B}_{-i}^\top(\boldsymbol{B}_{-i}^\top)^{+})\boldsymbol{\beta}_i\right|\right|_2 - \left|\left|\alpha_i \boldsymbol{B}_{-i}^\top \frac{\boldsymbol{\alpha}_{-i}}{||\boldsymbol{\alpha}_{-i}||_2^2}\right|\right|_2 \\
&\geq \left|\left|(\boldsymbol{I}_K - \boldsymbol{B}_{-i}^\top(\boldsymbol{B}_{-i}^\top)^{+})\boldsymbol{\beta}_i\right|\right|_2 - |\alpha_i| \frac{||\boldsymbol{B}_{-i}^\top||_2}{||\boldsymbol{\alpha}_{-i}||_2} \\
&\geq ||\boldsymbol{\beta}_i||_2 - |\alpha_i| \frac{||\boldsymbol{B}_{-i}^\top||_2}{||\boldsymbol{\alpha}_{-i}||_2} \\
&> 0,
\end{aligned}$$

where the first inequality is the reverse triangle inequality, and the second inequality follows from the sub-multiplicative property of the operator norm induced by the Euclidean norm. The third inequality uses the special properties of the matrix $\boldsymbol{I}_K - \boldsymbol{B}_{-i}^\top(\boldsymbol{B}_{-i}^\top)^{+}$. This matrix is an orthogonal projection matrix, and moreover orthogonal projection matrices have only two possible eigenvalues, 0 and 1. Moreover for a general square matrix A and a vector x, we have the inequality

$$||Ax||_2 \geq (\max_i |\lambda_i|)\, ||x||_2 \,,$$

where λ_i's are the eigenvalues of the matrix A. Apply this to our projection matrix and the maximum eigenvalue will simply be 1, and thus the third inequality follows. The last strict inequality is Assumption B.1(iii). Given that the norm of the second element of (28) is strictly positive, then the term itself cannot be zero. This shows that (28) is non-zero. □

Proof of Corollary 1.2. This is just a rewriting of the identification of $\boldsymbol{\Phi}_i$ from (23) in the proof of Theorem 1.1 as a $K+1$ vector, and where we denote,

$$\boldsymbol{A}_{-i}^{+} =: \boldsymbol{C}_i = \begin{bmatrix}\boldsymbol{c}_{i,0}^\top \\ \boldsymbol{c}_{i,1}^\top \\ \vdots \\ \boldsymbol{c}_{i,K}^\top\end{bmatrix} = \begin{bmatrix}\boldsymbol{\alpha}_{-i}^{+} \\ \boldsymbol{c}_{i,1}^\top \\ \vdots \\ \boldsymbol{c}_{i,K}^\top\end{bmatrix}$$

□

For simplicity, in the proof of Corollary 1.4 we assume that: $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i\varepsilon_j] = 0$ when $i \neq j$, $\mathbb{E}[\varepsilon_i^2] = \sigma_\varepsilon^2$, $\mathbb{E}[\boldsymbol{\Phi}_i\varepsilon_i] = \mathbf{0}_{K+1}$, and $\mathbb{E}[\boldsymbol{f}\varepsilon_i] = \mathbf{0}_{K+1}$. Furthermore, we assume that the variance-covariance matrices $\mathrm{Var}(\boldsymbol{\Phi}_i)$ and $\mathrm{Var}(\boldsymbol{f})$ are both positive definite.

Proof of Corollary 1.4. From (1b), we compute,

$$\begin{aligned}
\mathbb{E}[(R_i - \mathbf{b}_i^\top \boldsymbol{R}_{-i})^2] &= \mathrm{Var}(R_i - \mathbf{b}_i^\top \boldsymbol{R}_{-i}) + (\mathbb{E}[R_i - \mathbf{b}_i^\top \boldsymbol{R}_{-i}])^2 \\
&= \mathrm{Var}(\mathbf{a}_i^\top \boldsymbol{\Phi}_i - \mathbf{b}_i^\top \boldsymbol{\varepsilon}_{-i} + \varepsilon_i) + (\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i])^2 \\
&= \mathbf{a}_i^\top \mathrm{Var}(\boldsymbol{\Phi}_i)\mathbf{a}_i + (\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i])^2 + \sigma_\varepsilon^2 \mathbf{b}_i^\top \mathbf{b}_i + \sigma_\varepsilon^2,
\end{aligned}$$

and from (1a),

$$\mathbb{E}[(R_i - \mathbb{E}[R_i])^2] = \mathrm{Var}(R_i) = \mathbf{a}_i^\top \mathrm{Var}(\boldsymbol{f})\mathbf{a}_i + \sigma_\varepsilon^2.$$

Then we have that,

$$\begin{aligned}
\mathsf{R^2}_i &:= 1 - \frac{\mathbb{E}[(R_i - \mathbf{b}_i^\top \boldsymbol{R}_{-i})^2]}{\mathbb{E}[(R_i - \mathbb{E}[R_i])^2]} \\
&= 1 - \frac{\mathbf{a}_i^\top \mathrm{Var}(\boldsymbol{\Phi}_i)\mathbf{a}_i + (\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i])^2 + \sigma_\varepsilon^2(1 + \mathbf{b}_i^\top \mathbf{b}_i)}{\mathbf{a}_i^\top \mathrm{Var}(\boldsymbol{f})\mathbf{a}_i + \sigma_\varepsilon^2}
\end{aligned}$$

Part (a) follows because $\mathsf{R^2}_i$ depends on the magnitude of $\mathbf{a}_i^\top \mathrm{Var}(\boldsymbol{\Phi}_i)\mathbf{a}_i/(\mathbf{a}_i^\top \mathrm{Var}(\boldsymbol{f})\mathbf{a}_i + \sigma_\varepsilon^2)$. This is equivalent to considering the sign of $\mathbf{a}_i^\top(\mathrm{Var}(\boldsymbol{\Phi}_i) - \mathrm{Var}(\boldsymbol{f}))\mathbf{a}_i - \sigma_\varepsilon^2$. Since the variance-covariance matrix $\mathrm{Var}(\boldsymbol{f})$ is positive-definite, part (b) follows from the negative linear relationship between $\mathsf{R^2}_i$ and $(\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i])^2 = |\mathbf{a}_i^\top \mathbb{E}[\boldsymbol{\Phi}_i]|^2$. $\square$

B.2 Dependent bivariate portfolio sorts

In this section we investigate deeper into the nature of asset synchronicity and asset insurance premium by dependent bivariate portfolio sorts with other stock characteristics. We first sort the stocks i according to some characteristic of stock i into five equally sized bins. Then within each characteristic quantile bin, we further sort the stocks into 10 equally sized deciles according to their elastic-net R^2. Our dependent bivariate sort results thus evaluates the effects of elastic-net R^2, while directly controlling for the effects of the said characteristic. Moreover, in accordance to Theorem 1.1 and the main result of Section 3, we are looking for the *cross-sectional* effect of asset insurance premium and not just its existence in a given bivariate sorted bin. In other words for all the subsequent tables, we are most interested in the statistical results of the "Hi - Lo" column for the subtable labeled "Act - Gho". For brevity, we will only present value-weighted results in this section. The equal-weighted results are effectively identical to the value-weighted results here, and the equal-weighted results are available upon request.

B.2.1 Controlling for idiosyncratic related effects

By the conventional definition of R^2 as in (6), it will be negatively related to the variance of the returns of actual stock i over the estimation period. Thus it is important that the effects of the elastic-net R^2 are still present once we control for both the *total volatility* and the *idiosyncratic volatility* of stock i. In addition, we saw in Section 5 that the R^2's from the FF3 and FF5 regressions on individual stocks are highly correlated with the R^2's out of the elastic-net method. Even though FF3 and FF5 R^2's may not be a documented price factor in the literature, and indeed our results in Section 5 show there is no cross-sectional difference in returns of actual stocks across FFz R^2's, nonetheless it is prudent to control for their effects.

We define *total volatility* of actual stock i at month $t-1$ as the sample standard deviation of actual stock i's past twelve months of daily trading returns up until month $t-1$. *Idiosyncratic volatility* is as defined in Ang et al. (2006); we use OLS to regress the actual stock i's past twelve months of daily returns leading up until month $t-1$ onto the daily Fama-French three factors, and the resulting sample standard deviation of the residuals is the idiosyncratic volatility.

Table 25 shows the results that control for total volatility. The asset insurance premium in the cross-section of asset synchronicity is unaffected by a stock's total volatility. Stocks with the lowest [highest] total volatility have a cross-sectional asset insurance premium of -0.432% (t-stat -2.842) [-0.851% (t-stat -3.599)] per month, and average stocks have a cross-sectional premium of -0.851% (t-stat -3.599).

Table 26 shows the results that control for idiosyncratic volatility. Stocks with the lowest [highest] idiosyncratic volatility have a cross-sectional asset insurance premium of -0.481% (t-stat -3.274) [-0.951% (t-stat -2.302)] per month, whereas average stocks have a cross-sectional premium of -0.916% (t-stat -4.138).

One interpretation of the elastic-net R^2 is the goodness-of-fit of a stock i's return by the returns of all other stocks, and thus one might naturally regard a stock i with high R^2 to be associated with low idiosyncratic risk, and low R^2 with high idiosyncratic risk. Given the results of Ang et al. (2006), Table 26 is particularly important because it provides evidence that our results are showing something that idiosyncratic volatility is not.

[Table 25 about here.]

[Table 26 about here.]

Table 27 shows the results after controlling for FF3 R^2. Asset insurance premium effect persists in the cross-section of elastic-net defined asset synchronicity, even though its effects are statistically weaker for stocks with the highest FF3 R^2. For stocks with the lowest FF3 R^2 the cross-sectional asset insurance premium is -0.589% (t-stat -3.073) per month, but in the highest FF3 R^2 the cross-sectional premium is

-0.295% (t-stat -1.974), but on average the cross-sectional premium is -0.588% (t-stat -5.296). Table 28 shows the results after controlling for FF5 R^2. For stocks with the lowest FF5 R^2 the cross-sectional asset insurance premium is -0.881% (t-stat -4.535) per month, but in the highest FF5 R^2 the cross-sectional premium is statistically insignificant at -0.275% (t-stat -1.840). On average, after controlling for the effects of FF5 R^2, the cross-sectional premium is -0.611% (t-stat -5.749). Despite the weak or lack of statistical evidence of asset insurance premium for stocks with the highest FF3 and FF5 R^2, the fact that we find the premium in low to medium FF3 and FF5 R^2 stocks is still a positive result, especially since FFz R^2's are so positively correlate with elastic-net R^2's.

[Table 27 about here.]

[Table 28 about here.]

B.2.2 Controlling for corporate fundamentals

The size and value effects are one of the most robust and well-documented in the empirical asset pricing literature. It is thus important to show that the asset insurance premium is not driven by these two factors.

Table 29 shows asset insurance premia exist even controlling for firm size. The smallest [biggest] stocks have a cross-sectional asset insurance premium of -0.885% (t-stat -4.305) [-0.373% (t-stat -2.380)] per month. On average, the cross-sectional asset insurance premium after controlling for size is -0.745% (t-stat -4.528) per month. These bivariate result suggests small stocks demand an even higher asset insurance premia than large stocks.

[Table 29 about here.]

We investigate how the asset synchronicity effect is affects value and growth firms. Following conventional practice in the literature, we compute the book-to-market ratio breakpoints at the end of each June. Stocks of the lowest book-to-market ratios are "growth" stocks, while stocks of the highest book-to-market ratios are "value" stocks. The book value of equity used in June of a year is the book equity for the last fiscal year end of the previous year. Market equity is share prices multiplied by shares outstanding at the end of the December of the last fiscal year end of the previous year.

Table 30 shows the result and that asset synchronicity affects both value and growth stocks. Growth stocks have a cross-sectional asset insurance premium of -0.664% (t-stat -2.724) per month, value stocks have a cross-sectional premium of -0.791% (t-stat -2.721). The average cross-sectional insurance premium after controlling for book-to-market effects is -0.606% (t-stat -3.134).

[Table 30 about here.]

B.2.3 Controlling for liquidity

The data filtering and estimation procedure of Section 2 could actually include numerous illiquid stocks. We need to ensure that our asset insurance premium effect is not simply reflecting a liquidity premium or an illiquidity discount.

Table 31 shows the effects of asset insurance premium after controlling for dollar volume liquidity. We see that stocks with low [high] dollar volume liquidity have a cross-sectional asset insurance premium of -0.480% (t-stat -2.681) [-0.415% (t-stat -2.654)] per month. Across average stocks after controlling for dollar volume liquidity, the cross-sectional asset insurance premium is -0.506% (t-stat -3.130) per month.

[Table 31 about here.]

Table 32 shows the effects of Amihud (2002)'s illiquidity measure on asset insurance premium. We find stocks with low illiquidity (so liquid stocks) have a cross-sectional asset insurance premium of -0.398% (t-stat -2.617) per month. The statistical evidence for a cross-sectional asset insurance premium is weak for stocks with high illiquidity, where we see the estimate is -0.457% (t-stat -1.842) per month. But on average, controlling for the effects of illiquidity, we still find a cross-sectional asset insurance premium of -0.559% (t-stat -3.134) per month.

[Table 32 about here.]

B.2.4 Controlling for factors based on past returns

Our entire estimation and portfolio construction method of Section 2 depends on using past data. An important robustness check for our results is to control for known priced factors that are also based on past data.

Jegadeesh (1990) and Lehmann (1990) document that short-term reversal is a priced factor in the cross-section. Table 33 shows that stocks with the lowest [highest] short-term reversal earn an asset insurance premium of -1.188% (t-stat -3.515) [-0.678% (t-stat -2.395)] per month. The average asset insurance premium after controlling for short-term reversal is -0.730% (t-stat -3.396) per month.

[Table 33 about here.]

Momentum is known to be a challenging priced factor to control. For our case, we see that stocks with the lowest momentum have a statistically insignificant asset insurance premium of -0.174% (t-stat -0.509) per month. In contrast, stocks with the highest momentum have a economically large asset insurance premium of -169.2% (t-stat -4.775) per month. On average, the asset insurance premium after controlling for momentum is -0.685% (t-stat -3.251) per month.

[Table 34 about here.]

Figure 1: Projection procedure. For each month ending at December 31, 1975, January 31, 1976, ..., December 31, 2017, we use the past twelve months' worth of daily observations to project each of stock returns onto the returns span of every other stock. There are $T = 504$ number of months from December 31, 1975 to December 31, 2017. We only consider stocks have at least 60 days worth for daily trading data. That is, for each stock $i = 1, \ldots, N_{i,t-1}$, suppose $\{d_1, \ldots, D_{i,t-1}\}$ with $D_i \geq 60$ are past twelve months' worth of trading days that end at month $t-1$. Note that the number of stocks $N_{i,t-1}$ available per month end $t-1$ may vary. The returns span for stock i are all those other stocks $j = 1, \ldots, i-1, i+1, \ldots, N_{i,t-1}$ that have trading days that at least overlap with that of stock i, so $\{d_1, \ldots, D_{i,t-1}\} \subseteq \{d_1, \ldots, D_{j,t-1}\}$.

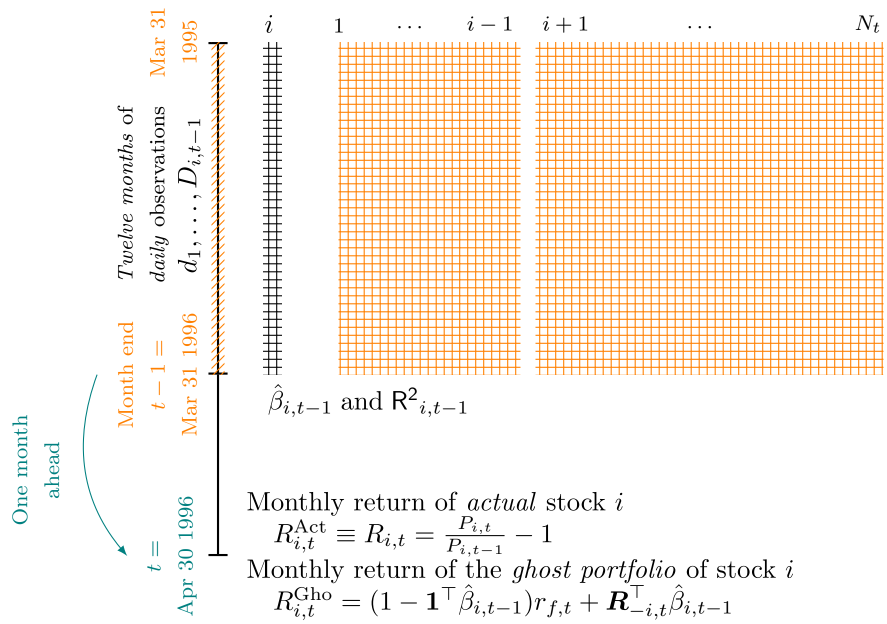

Figure 2: Dimensionality of the returns span for each stock at each month end. For month ends t = December 31, 1975, January 31, 1976, ..., December 31, 2017, we use past twelve months worth of daily observations ending at month $t-1$ to project each of stock i's returns onto the returns of *all other* stocks. We only consider stocks that have at least 60 trading days with non-zero returns ending at month $t-1$. Other than the additional overlapping return filter that we'd described in Section 2, this time series thus effectively plots the dimensionality $N_{i,t-1}$ of the asset span for each actual stock i across time.

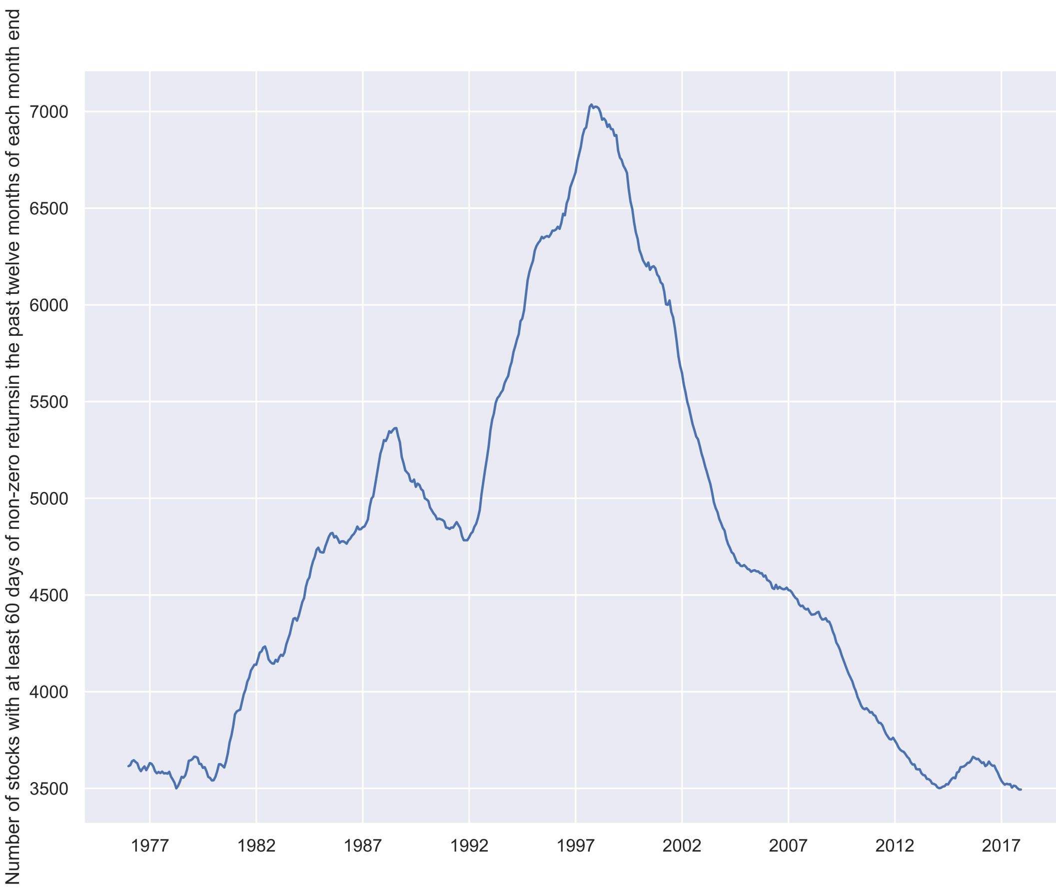

Figure 3: Three types of returns for a given risky asset. For any given stock i, we consider three types of returns: (i) its *actual returns* $R_i^{\mathrm{Act}} = R_i$; (ii) its *ghost (portfolio of) returns* R_i^{Gho}, which represents the closest projected representation of the actual stock i return in the span of all other stocks in the market; and (iii) its *long-short actuals versus ghost returns* R_i^{Ls}, which can be thought of as the projection residual between the actual stock return versus its projected representation in the span of all other stocks. The term $R_{i,t}^{\mathrm{Gho}}$ will empirically proxy for $\mathbf{b}_i^\top \boldsymbol{R}_{-i}$ in Theorem 1.1, while the term $R_{i,t}^{\mathrm{Ls}}$ will proxy for $R_i - \mathbf{b}_i^\top \boldsymbol{R}_{-i}$.

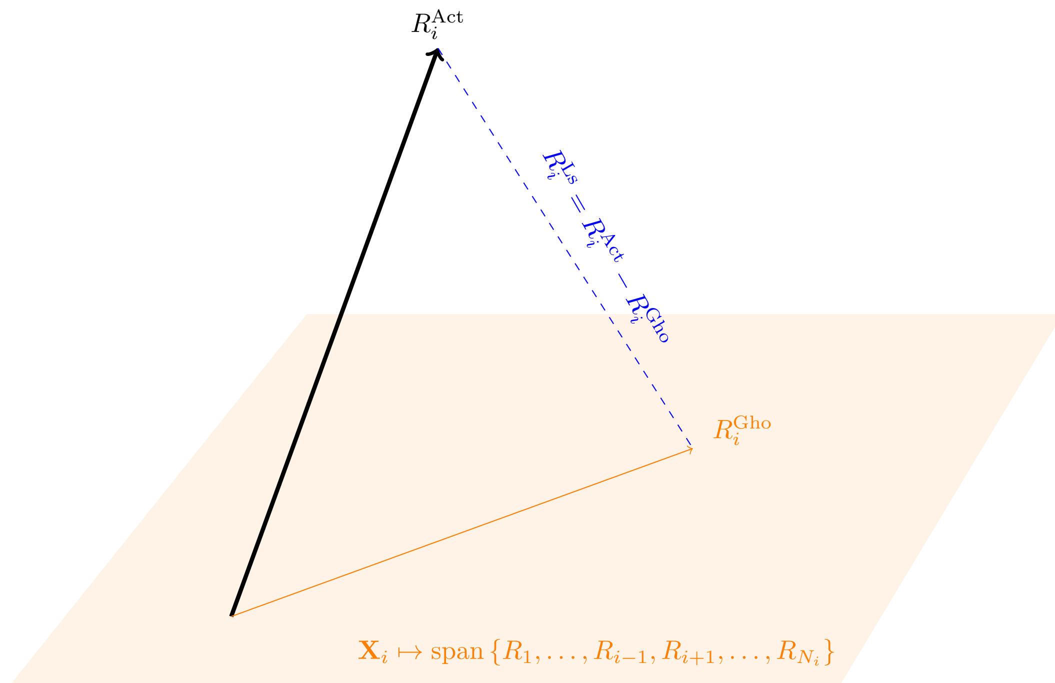

Figure 4: Time series of the cross-sectional R^2 averages of the elastic-net, FF3 and FF5 procedures At the end of month $t-1$, we compute for each stock i the $\mathsf{R}^2{}_{i,t-1}$ coefficients under all of the elastic-net estimation method of Section 2, and also under the FFz methods of Section 5. We then compute the simple *cross-sectional average* of all the $\mathsf{R}^2{}_{i,t-1}$ across i's; that is, we compute $\bar{\mathsf{R}}^2_{t-1} := \frac{1}{N_{t-1}} \sum_{i=1}^{N_{t-1}} \mathsf{R}^2{}_{i,t-1}$. Here, N_{t-1} is the total number of traded stocks at month $t-1$, subject to the minimum trading days requirement outlined in Section 2. We also plot the differenced quantity, $\Delta\bar{\mathsf{R}}^2_t := \bar{\mathsf{R}}^2_t - \bar{\mathsf{R}}^2_{t-1}$. The red shaded bars are NBER recession dates. The sample observation is from December 1975 to December 2017.

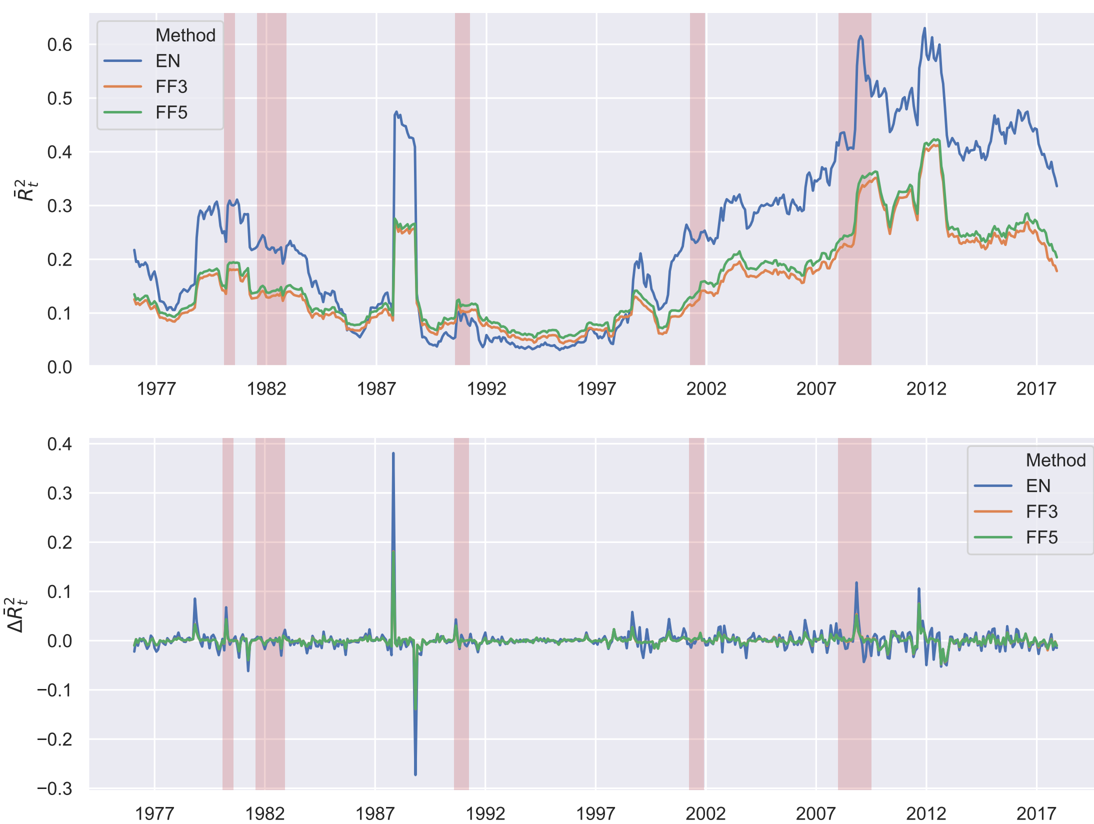

Figure 5: Average proportion of equity invested in the ghost portfolios by selected elastic-net, FF3 and FF5 R^2 bins. Consider the average amount of equity-only components that are invested by each ghost portfolio. Recall the construction of the ghost portfolio in (8) of the elastic-net method, and (16) for FF3 and FF5. At the end of the estimation month $t-1$, we will have the vector $\hat{\beta}_{m,i,t-1}$ for method $m \in \{\mathrm{EN}, \mathrm{FF3}, \mathrm{FF5}\}$ of asset i. The quantity $1 - \mathbf{1}^\top \hat{\beta}_{m,i,t-1}$ is the proportion of wealth allocated to the risk free rate in the ghost portfolio of stock i, while $\mathbf{1}^\top \hat{\beta}_{m,i,t-1}$ is the proportion allocated to equity-only components. The stocks are sorted into their method m R^2 decile bins $k = \text{'Lo'}, \ldots, \text{'Hi'}$. The average amount of equity-only components that are invested in each ghost portfolio at month $t-1$ of method m is the quantity $\frac{1}{|B^k_{m,t-1}|} \sum_{i \in B^k_{m,t-1}} \mathbf{1}^\top \hat{\beta}_{m,i,t-1}$. The above plots the bins $k = \text{'Lo'}, 5$ and 'Hi', and the y-axis are expressed in decimals (e.g. 0.10 means 10%). The plot labeled "Overall average" plots the overall average of these equity-only components, which is the quantity $\frac{1}{10} \sum_{k=1}^{10} \frac{1}{|B^k_{m,t-1}|} \sum_{i \in B^k_{m,t-1}} \mathbf{1}^\top \hat{\beta}_{m,i,t-1}$. The red shaded bars are NBER recession dates. The sample is from December 1975 to December 2017.

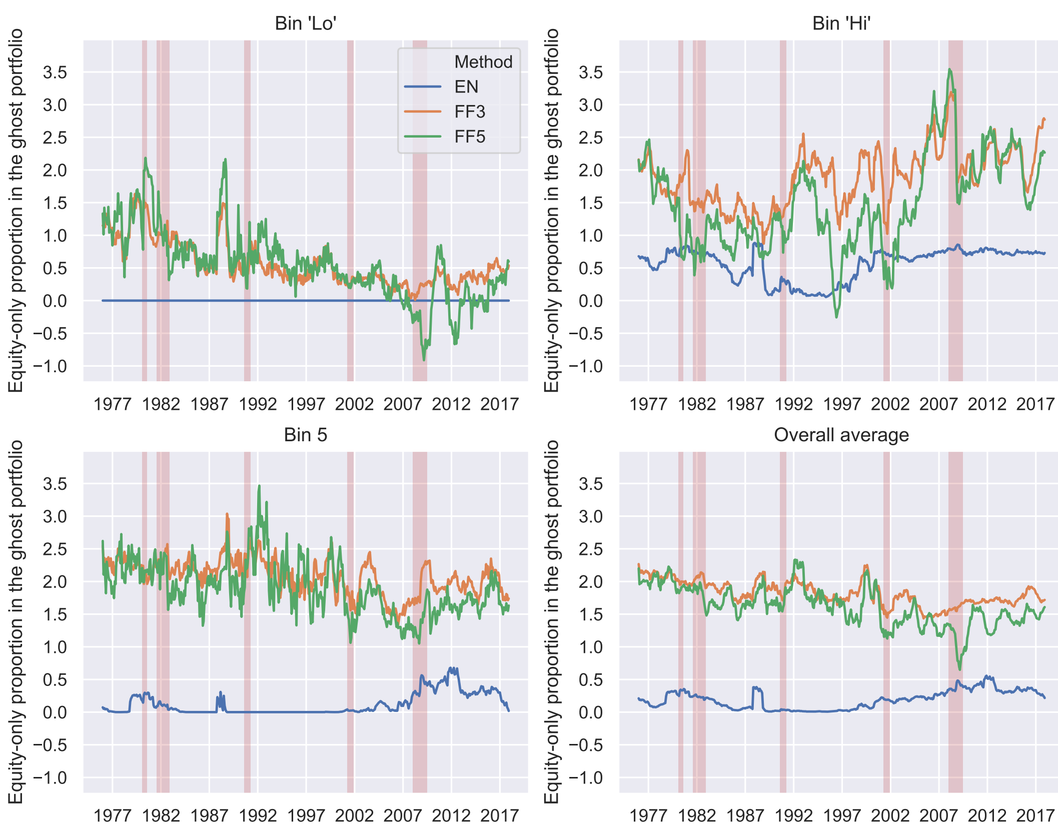

Table 1: Summary statistics of the estimated elastic-net coefficient vector. Actual stocks i are sorted by their elastic-net $\mathsf{R}^2{}_{i,t-1}$ at month $t-1$ into bins $k = \text{Lo}, 2, ..., \text{Hi}$ in ascending order. Each kth bin consists of a panel of estimated elastic-net coefficient vectors $\hat{\beta}_{i,t-1}$. Table (a) reports the summary statistics on the number of non-zero elements in the estimated coefficient vectors, not conditioning on time nor the cross-section. For instance in Table (a), the mean of the kth bin is calculated as $\frac{1}{T}\sum_t \frac{1}{|B_t^k|}\sum_{i\in B_t^k}(\#\text{ of non-zero elements in }\hat{\beta}_{i,t})$; other statistics are calculated analogously. Table (b) reports the summary statistics on the sum of elements in the estimated coefficient vectors, not conditioning on time nor the cross-section. For instance in Table (b), the mean of the kth bin is calculated as $\frac{1}{T}\sum_t \frac{1}{|B_t^k|}\sum_{i\in B_t^k} \mathbf{1}^\top\hat{\beta}_{i,t}$; other statistics are calculated analogously. Recall from Section 2.2 that for each actual stock i at month $t-1$, the ghost portfolio of stock i has $1 - \mathbf{1}^\top\hat{\beta}_{i,t}$ proportion of wealth invested into the risk free asset. The columns "p5", "p25", "med", "p75" and "p95" respectively represent the 5th, 25th, 50th, 75th and 95th percentiles of the data. The sample period is from December 1975 to December 2017.

EN R^2	mean	std	min	max	p5	p25	med	p75	p95
Lo	0.505	0.594	0	10	0	0	0	1	1
2	0.639	0.970	0	22	0	0	1	1	2
3	2.004	4.512	0	58	0	0	1	1	11
4	5.448	10.087	0	91	0	0	1	5	29
5	10.184	15.104	0	116	0	1	2	16	43
6	16.059	19.701	0	132	0	1	6	28	56
7	23.024	24.396	0	152	1	2	14	39	70
8	31.407	29.591	1	167	1	4	26	51	87
9	43.724	35.402	1	205	3	9	40	67	109
Hi	79.032	52.403	1	448	10	41	71	108	177

(a) Summary statistics on the count of non-zero elements in the estimated elastic-net coefficient vector in each elastic-net R^2 bin

EN R^2	mean	std	min	max	p5	p25	med	p75	p95
Lo	0.000	0.002	-0.147	0.538	-0.000	0.000	0.000	0.000	0.000
2	0.001	0.009	-0.132	0.427	-0.000	0.000	0.000	0.000	0.006
3	0.017	0.063	-0.265	1.078	-0.000	0.000	0.000	0.000	0.109
4	0.055	0.132	-0.485	1.658	-0.000	0.000	0.000	0.033	0.340
5	0.102	0.178	-0.342	1.771	-0.000	0.000	0.006	0.139	0.494
6	0.156	0.219	-0.472	2.117	-0.001	0.000	0.034	0.263	0.619
7	0.214	0.256	-0.216	2.245	-0.004	0.003	0.106	0.374	0.721
8	0.278	0.289	-0.399	2.580	-0.006	0.010	0.210	0.483	0.815
9	0.364	0.320	-0.293	3.289	-0.006	0.033	0.346	0.603	0.915
Hi	0.531	0.337	-0.391	4.674	0.008	0.261	0.567	0.773	1.043

(b) Summary statistics on sum of elements of the estimated elastic-net coefficient vector by elastic-net R^2 bins

Table 2: Summary statistics of characteristics of actual stocks sorted by elastic-net R^2. Actual stocks i are sorted by their elastic-net $\mathsf{R}^2{}_{i,t-1}$ at month $t-1$ into bins $\text{Lo}, 2, \ldots, \text{Hi}$ in ascending order. Each kth bin consists of a panel of characteristics of stock i at month $t-1$. The above reports the summary statistics on the panel of characteristics of the actual stocks that is not conditioned by time nor the cross-section. "MktCap" is the market capitalization of the actual stocks, which is defined as the share price multiplied by shares outstanding, divided by 1,000,000. "B/M" is the book-to-market ratio of actual stocks, where we merge the book value data from Compustat by conventional practices. "VOLD" is dollar trading volume, which is defined as share price multiplied by volume, divided by 1,000,000. "TotalVol" is the standard deviation (in decimals) of one year's worth of daily returns up to month $t-1$ of the actual stock. "EN R^2" is the level of the R^2 of actual stock i at month $t-1$ from the procedure of Section 2. "FF5 R^2" is the level of the R^2 resulting from regressing actual stock i's one year's worth of daily returns up to month $t-1$ onto the daily Fama-French five factors (15). The columns "p5", "p25", "med", "p75" and "p95" respectively represent the 5th, 25th, 50th, 75th and 95th percentiles of the data. The sample period is from December 1975 to December 2017.

	MktCap									B/M								
EN R^2	mean	std	min	max	p5	p25	med	p75	p95	mean	std	min	max	p5	p25	med	p75	p95
Lo	1029.017	7621.841	0.071	867904.049	5.393	26.633	82.276	298.924	3257.307	0.964	1.138	0.000	53.135	0.120	0.393	0.730	1.200	2.467
2	769.916	6441.566	0.028	697505.421	4.200	17.813	53.486	178.556	2121.897	0.991	1.156	0.000	53.135	0.121	0.401	0.744	1.230	2.576
3	771.615	7224.570	0.046	643120.129	3.875	16.253	49.765	161.395	1899.880	0.970	1.161	0.000	53.135	0.115	0.381	0.717	1.198	2.552
4	1092.241	9980.904	0.044	748247.186	4.007	17.372	59.479	223.053	2508.551	0.915	1.137	0.000	81.304	0.107	0.352	0.670	1.136	2.387
5	1385.725	10874.820	0.062	847097.392	4.435	20.292	79.472	336.581	3583.480	0.866	1.047	0.000	81.304	0.103	0.332	0.635	1.088	2.254
6	1583.358	11427.499	0.011	768224.360	5.103	24.897	112.231	490.894	4495.680	0.822	0.971	0.000	81.304	0.105	0.323	0.607	1.038	2.138
7	1840.414	11754.129	0.058	747867.383	6.238	33.994	165.888	701.165	5716.362	0.795	0.931	0.000	81.304	0.107	0.326	0.597	1.001	2.032
8	2180.792	12453.065	0.042	882331.569	7.455	43.056	217.305	927.486	7317.859	0.763	0.882	0.000	81.304	0.107	0.325	0.582	0.963	1.909
9	2828.560	13643.401	0.101	629010.237	9.069	58.976	315.511	1324.763	10356.478	0.731	0.844	0.000	81.304	0.103	0.319	0.565	0.924	1.814
Hi	5274.532	20445.302	0.234	602432.919	14.160	129.432	712.868	2884.035	21386.085	0.679	0.676	0.000	40.578	0.102	0.308	0.537	0.869	1.661

	VOLD									TotalVol								
EN R^2	mean	std	min	max	p5	p25	med	p75	p95	mean	std	min	max	p5	p25	med	p75	p95
Lo	5.623	49.634	0.000	6225.065	0.013	0.068	0.279	1.464	18.805	0.037	0.038	0.002	1.247	0.010	0.018	0.028	0.044	0.091
2	3.878	32.338	0.000	5804.049	0.008	0.037	0.135	0.712	11.717	0.040	0.032	0.002	0.884	0.012	0.020	0.031	0.049	0.098
3	4.415	41.872	0.000	5864.435	0.007	0.035	0.126	0.664	11.965	0.040	0.028	0.002	0.656	0.013	0.022	0.033	0.050	0.093
4	7.112	76.686	0.000	10143.954	0.008	0.042	0.178	1.094	20.549	0.039	0.025	0.001	0.416	0.013	0.022	0.033	0.049	0.088
5	9.995	97.336	0.000	11771.581	0.009	0.052	0.278	1.945	31.902	0.038	0.024	0.001	0.404	0.013	0.021	0.032	0.047	0.084
6	11.689	94.182	0.000	11299.650	0.011	0.072	0.434	3.080	41.143	0.036	0.023	0.001	0.605	0.013	0.021	0.030	0.045	0.081
7	13.574	87.318	0.000	6322.208	0.014	0.105	0.657	4.456	52.471	0.035	0.023	0.004	0.807	0.012	0.020	0.029	0.043	0.079
8	16.255	86.694	0.000	6133.136	0.017	0.138	0.892	6.133	69.254	0.034	0.023	0.003	0.461	0.012	0.019	0.028	0.042	0.079
9	22.127	95.197	0.000	6330.419	0.020	0.197	1.391	9.603	103.301	0.034	0.023	0.003	0.545	0.012	0.019	0.027	0.041	0.078
Hi	39.194	141.094	0.000	3907.545	0.031	0.489	3.727	21.127	178.361	0.031	0.021	0.003	0.802	0.012	0.018	0.026	0.037	0.070

	EN R^2									FF5 R^2								
EN R^2	mean	std	min	max	p5	p25	med	p75	p95	mean	std	min	max	p5	p25	med	p75	p95
Lo	-0.012	0.010	-0.307	0.004	-0.029	-0.015	-0.009	-0.005	-0.002	0.074	0.069	0.000	0.707	0.010	0.027	0.052	0.099	0.211
2	-0.002	0.009	-0.019	0.146	-0.009	-0.005	-0.003	-0.001	0.008	0.066	0.065	0.000	0.674	0.009	0.023	0.044	0.086	0.198
3	0.017	0.052	-0.010	0.431	-0.005	-0.002	-0.001	0.001	0.125	0.068	0.065	0.000	0.660	0.009	0.023	0.046	0.090	0.200
4	0.058	0.115	-0.005	0.622	-0.002	-0.001	-0.000	0.052	0.339	0.084	0.083	0.000	0.671	0.009	0.026	0.054	0.115	0.260
5	0.110	0.166	-0.002	0.721	-0.001	-0.000	0.016	0.175	0.489	0.108	0.106	0.000	0.736	0.010	0.030	0.069	0.154	0.339
6	0.169	0.206	-0.001	0.790	-0.000	-0.000	0.066	0.298	0.593	0.135	0.126	0.000	0.765	0.011	0.036	0.094	0.199	0.402
7	0.233	0.241	-0.000	0.851	0.001	0.009	0.162	0.409	0.681	0.165	0.144	0.000	0.813	0.013	0.045	0.125	0.248	0.458
8	0.305	0.269	0.006	0.904	0.012	0.030	0.283	0.534	0.770	0.195	0.163	0.000	0.813	0.014	0.054	0.157	0.298	0.518
9	0.401	0.287	0.020	0.955	0.034	0.086	0.414	0.662	0.850	0.234	0.183	0.000	0.844	0.015	0.069	0.202	0.356	0.584
Hi	0.606	0.281	0.054	1.000	0.103	0.400	0.656	0.849	0.968	0.302	0.202	0.000	0.895	0.020	0.132	0.284	0.445	0.667

Table 3: **(MAIN RESULTS)** **Univariate equal- and value- weighted portfolio sort by elastic-net R^2.** We form equal- and value-weighted portfolios decile portfolios every month by regressing each stock's daily return over the past year onto every other stock using the elastic-net estimator (24). Stocks are sorted into deciles based on their elastic-net R^2 from the lowest (quantile 1, labelled "Lo") to highest (quantile 10, labelled "Hi"). The row labelled "Hi - Lo" is the monthly return difference between the "Hi" bin and the "Lo" bin. The row labelled "Avg" is the simple average of the monthly returns across the ten $k = \text{'Lo'}, 2, \ldots, \text{'Hi'}$ bins. The column labelled "Actual" reports the one month ahead portfolio returns of actual stocks (7). The column labelled "Ghost" reports the one month ahead returns of a portfolio that are constructed out of the estimated elastic-net beta coefficients according to (8). The column labelled "Act - Gho" reports the one month ahead returns of the portfolio of a long position in the actual stocks, and a short position in the corresponding ghost portfolios of the actual stocks. The value weight of the ghost portfolio of stock i is the same value weight as the actual stock i. For "Actual" and "Ghost", the columns labelled "mean" and "sd" are excess returns measured in monthly percentage terms (e.g. 1.0 means 1%). For "Act - Gho", the columns labelled "mean" and "sd" are total simple returns, and not excess returns, measured in monthly percentage terms. Robust Newey and West (1987) t-statistics are reported in column "t" in parentheses. The sample period is from December 1975 to December 2017.

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
EN R^2									
Lo	1.440	4.964	(5.024)	-0.000	0.527	(-0.204)	1.440	4.966	(5.027)
2	1.158	4.877	(3.962)	0.005	0.661	(1.273)	1.153	4.872	(3.957)
3	0.958	4.980	(3.167)	0.043	1.662	(1.444)	0.916	4.890	(3.141)
4	0.889	4.972	(2.971)	0.104	2.606	(1.289)	0.788	4.717	(2.948)
5	0.875	5.014	(3.006)	0.136	3.082	(1.176)	0.740	4.571	(3.085)
6	0.849	5.109	(2.976)	0.153	3.425	(1.102)	0.698	4.483	(3.244)
7	0.860	5.179	(3.039)	0.223	3.786	(1.360)	0.639	4.308	(3.285)
8	0.890	5.333	(3.099)	0.254	4.219	(1.309)	0.640	4.152	(3.617)
9	0.824	5.464	(2.876)	0.280	4.669	(1.249)	0.546	3.972	(3.541)
Hi	0.786	5.360	(2.827)	0.416	4.898	(1.670)	0.371	3.384	(3.245)
Hi - Lo	-0.653	4.011	(-4.246)	0.417	4.899	(1.674)	-1.069	4.684	(-3.904)
Avg	0.953	5.023	(3.406)	0.161	3.170	(1.437)	0.793	4.256	(3.942)

(a) Equal-weighted portfolios

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
EN R^2									
Lo	0.969	4.366	(5.034)	-0.001	0.529	(-0.418)	0.976	4.368	(5.081)
2	0.817	4.311	(4.341)	0.003	0.698	(0.819)	0.817	4.306	(4.364)
3	0.581	4.220	(3.117)	0.028	1.417	(1.358)	0.557	4.176	(3.057)
4	0.573	4.183	(3.176)	0.055	1.970	(1.235)	0.518	4.097	(3.089)
5	0.621	4.123	(3.672)	0.059	2.302	(0.969)	0.564	3.933	(3.792)
6	0.542	4.250	(2.981)	0.095	2.734	(1.088)	0.448	3.930	(2.950)
7	0.751	4.374	(3.863)	0.144	3.071	(1.295)	0.608	3.925	(4.156)
8	0.640	4.491	(3.155)	0.163	3.440	(1.192)	0.480	3.845	(3.460)
9	0.639	4.656	(2.842)	0.158	3.907	(0.899)	0.482	3.654	(3.825)
Hi	0.628	4.795	(2.677)	0.325	4.314	(1.617)	0.305	3.342	(2.700)
Hi - Lo	-0.341	3.954	(-2.174)	0.326	4.314	(1.623)	-0.671	4.138	(-3.536)
Avg	0.676	4.191	(3.757)	0.103	2.613	(1.307)	0.575	3.705	(4.333)

(b) Value-weighted portfolios

Table 4: Pairwise correlations between the elastic-net R^2 sorted portfolio bins of actual stocks against their ghost stocks. The (k, l)-th entry above represents the correlation of the time series one month ahead returns between the k-th portfolio bin of actual stocks and the l-th portfolio bin of ghost stocks. The entries are expressed in decimals (e.g. 0.10 means 10%). We form equal- and value-weighted portfolios decile portfolios every month by regressing each stock's daily return over the past year onto every other stock using the elastic-net estimator (24). Stocks are sorted into deciles based on their elastic-net R^2 from the lowest (quantile 1, labelled "Lo") to highest (quantile 10, labelled "Hi"). The rows labelled "Actual" report the portfolio of actual stocks (7). The columns labelled "Ghost" report the portfolio of returns that are constructed out of the estimated elastic-net beta coefficients according to (8). The time sapmle is from December 1975 to December 2017.

EN R^2	Lo Ghost	2	3	4	5	6	7	8	9	Hi Ghost
Lo Actual	-0.096	0.099	0.319	0.418	0.494	0.563	0.614	0.638	0.663	0.775
2	-0.089	0.111	0.319	0.416	0.501	0.583	0.644	0.676	0.699	0.795
3	-0.091	0.142	0.374	0.466	0.547	0.627	0.686	0.711	0.724	0.812
4	-0.093	0.114	0.355	0.483	0.579	0.666	0.722	0.744	0.749	0.828
5	-0.088	0.118	0.355	0.493	0.598	0.691	0.750	0.772	0.777	0.853
6	-0.080	0.108	0.334	0.473	0.579	0.679	0.749	0.781	0.796	0.876
7	-0.060	0.118	0.329	0.463	0.568	0.673	0.755	0.803	0.829	0.903
8	-0.049	0.121	0.319	0.446	0.548	0.655	0.750	0.819	0.860	0.929
9	-0.053	0.101	0.288	0.406	0.505	0.614	0.715	0.795	0.858	0.938
Hi Actual	-0.043	0.105	0.285	0.402	0.497	0.596	0.680	0.743	0.802	0.921

(a) Equal-weighted portfolios

EN R^2	Lo Ghost	2	3	4	5	6	7	8	9	Hi Ghost
Lo Actual	-0.077	0.006	0.192	0.284	0.344	0.380	0.430	0.482	0.507	0.638
2	-0.056	0.109	0.246	0.353	0.420	0.480	0.523	0.560	0.575	0.661
3	-0.076	0.084	0.233	0.323	0.385	0.424	0.462	0.491	0.511	0.620
4	-0.082	0.010	0.162	0.291	0.364	0.417	0.460	0.486	0.503	0.612
5	-0.084	0.036	0.221	0.353	0.432	0.480	0.519	0.545	0.568	0.662
6	-0.049	0.084	0.246	0.385	0.468	0.533	0.560	0.589	0.609	0.708
7	-0.053	0.088	0.258	0.393	0.477	0.542	0.603	0.648	0.685	0.766
8	-0.043	0.115	0.306	0.441	0.518	0.580	0.626	0.688	0.730	0.811
9	-0.041	0.121	0.301	0.433	0.514	0.587	0.644	0.721	0.793	0.861
Hi Actual	-0.014	0.109	0.271	0.392	0.472	0.535	0.598	0.679	0.746	0.877

(b) Value-weighted portfolios

Table 5: Fama-French *three* factor OLS regression on *equal*-weighted portfolios sorted by elastic-net R^2. We run the Fama-French 3 factor time series regression (13) onto equal-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's elastic-net R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k = \text{'Lo'}, 2, \ldots, \text{'Hi'}$ portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
EN R^2		coef	sd	t	coef	sd	t	coef	sd	t
Lo	const	1.201	0.244	(4.915)	-0.000	0.002	(-0.118)	1.201	0.244	(4.921)
	Mkt-RF	28.586	6.897	(4.145)	-0.018	0.085	(-0.213)	28.604	6.875	(4.160)
	HML	0.550	8.118	(0.068)	-0.082	0.137	(-0.601)	0.633	8.109	(0.078)
	SMB	19.967	8.101	(2.465)	0.088	0.088	(0.998)	19.878	8.103	(2.453)
2	const	0.862	0.232	(3.710)	0.004	0.004	(1.019)	0.858	0.231	(3.705)
	Mkt-RF	34.686	5.852	(5.927)	0.094	0.108	(0.876)	34.594	5.811	(5.954)
	HML	4.717	7.454	(0.633)	0.058	0.198	(0.292)	4.665	7.439	(0.627)
	SMB	21.837	7.253	(3.011)	0.134	0.097	(1.386)	21.708	7.243	(2.997)
3	const	0.660	0.241	(2.740)	0.028	0.026	(1.081)	0.632	0.231	(2.735)
	Mkt-RF	32.646	6.406	(5.096)	1.119	0.862	(1.298)	31.561	5.924	(5.327)
	HML	7.182	8.444	(0.851)	1.802	1.265	(1.425)	5.411	8.250	(0.656)
	SMB	24.658	7.701	(3.202)	0.736	0.627	(1.173)	23.956	7.569	(3.165)
4	const	0.630	0.245	(2.571)	0.057	0.069	(0.819)	0.577	0.217	(2.660)
	Mkt-RF	31.739	6.067	(5.231)	4.474	2.671	(1.675)	27.203	4.878	(5.577)
	HML	1.045	8.390	(0.125)	5.011	3.542	(1.415)	-4.121	8.169	(-0.504)
	SMB	19.202	7.495	(2.562)	0.829	1.545	(0.537)	18.283	7.145	(2.559)
5	const	0.656	0.245	(2.677)	0.067	0.101	(0.663)	0.590	0.199	(2.957)
	Mkt-RF	27.340	6.530	(4.187)	7.786	3.978	(1.957)	19.612	4.462	(4.395)
	HML	0.005	8.473	(0.001)	5.822	4.720	(1.233)	-5.768	7.159	(-0.806)
	SMB	15.873	7.936	(2.000)	-0.088	2.336	(-0.038)	15.944	7.162	(2.226)
6	const	0.661	0.248	(2.663)	0.069	0.125	(0.550)	0.593	0.187	(3.171)
	Mkt-RF	27.171	6.380	(4.259)	10.727	4.729	(2.268)	16.477	4.376	(3.765)
	HML	-4.665	8.856	(-0.527)	5.542	5.398	(1.027)	-10.115	7.319	(-1.382)
	SMB	10.092	8.362	(1.207)	-1.401	2.931	(-0.478)	11.527	6.905	(1.669)
7	const	0.700	0.255	(2.740)	0.128	0.148	(0.863)	0.573	0.177	(3.235)
	Mkt-RF	25.320	6.991	(3.622)	13.471	5.332	(2.526)	11.914	4.242	(2.808)
	HML	-8.338	9.112	(-0.915)	4.479	6.365	(0.704)	-12.721	5.915	(-2.151)
	SMB	8.177	8.792	(0.930)	-2.796	3.842	(-0.728)	10.991	6.603	(1.665)
8	const	0.758	0.263	(2.876)	0.149	0.177	(0.841)	0.612	0.168	(3.646)
	Mkt-RF	23.634	7.294	(3.240)	16.635	5.937	(2.802)	7.067	4.179	(1.691)
	HML	-10.908	9.874	(-1.105)	2.259	7.699	(0.293)	-13.129	5.428	(-2.419)
	SMB	4.446	9.298	(0.478)	-4.329	4.934	(-0.877)	8.822	6.249	(1.412)
9	const	0.737	0.272	(2.703)	0.185	0.209	(0.886)	0.555	0.154	(3.605)
	Mkt-RF	18.282	7.670	(2.384)	17.867	6.694	(2.669)	0.398	3.898	(0.102)
	HML	-12.496	11.162	(-1.119)	-2.718	9.477	(-0.287)	-10.028	5.018	(-1.998)
	SMB	2.531	9.865	(0.257)	-5.416	6.599	(-0.821)	7.823	5.827	(1.343)
Hi	const	0.727	0.277	(2.624)	0.324	0.234	(1.384)	0.404	0.122	(3.301)
	Mkt-RF	12.644	7.981	(1.584)	16.663	7.299	(2.283)	-3.986	2.656	(-1.501)
	HML	-9.306	11.104	(-0.838)	-6.505	10.867	(-0.599)	-2.780	3.322	(-0.837)
	SMB	2.045	9.838	(0.208)	1.424	8.319	(0.171)	0.728	3.716	(0.196)
Hi - Lo	const	-0.474	0.158	(-3.002)	0.324	0.233	(1.388)	-0.797	0.225	(-3.540)
	Mkt-RF	-15.942	3.363	(-4.741)	16.681	7.275	(2.293)	-32.590	6.431	(-5.067)
	HML	-9.856	6.385	(-1.544)	-6.423	10.849	(-0.592)	-3.413	8.459	(-0.403)
	SMB	-17.922	5.850	(-3.064)	1.337	8.314	(0.161)	-19.151	7.476	(-2.562)
Avg	const	0.759	0.241	(3.147)	0.101	0.102	(0.992)	0.660	0.172	(3.832)
	Mkt-RF	26.205	6.542	(4.006)	8.882	3.624	(2.451)	17.344	3.992	(4.344)
	HML	-3.221	8.560	(-0.376)	1.567	4.573	(0.343)	-4.795	5.542	(-0.865)
	SMB	12.883	7.978	(1.615)	-1.082	2.836	(-0.382)	13.966	5.973	(2.338)

EN R^2		**Actual**	**Ghost**	**Act - Gho**
Lo	F	14.525	0.687	14.548
	R^2	0.073	0.003	0.072
	Adj-R^2	0.067	-0.003	0.067
	nobs	504	504	504
2	F	25.089	0.786	25.192
	R^2	0.107	0.004	0.106
	Adj-R^2	0.101	-0.002	0.101
	nobs	504	504	504
3	F	18.482	0.749	20.450
	R^2	0.093	0.012	0.094
	Adj-R^2	0.088	0.006	0.089
	nobs	504	504	504
4	F	17.536	0.967	19.342
	R^2	0.083	0.020	0.084
	Adj-R^2	0.078	0.014	0.078
	nobs	504	504	504
5	F	12.343	1.428	13.269
	R^2	0.060	0.025	0.057
	Adj-R^2	0.054	0.019	0.051
	nobs	504	504	504
6	F	11.357	2.054	10.652
	R^2	0.051	0.029	0.047
	Adj-R^2	0.046	0.023	0.041
	nobs	504	504	504
7	F	10.135	2.790	9.322
	R^2	0.044	0.030	0.041
	Adj-R^2	0.039	0.024	0.035
	nobs	504	504	504
8	F	8.057	3.742	5.736
	R^2	0.035	0.030	0.028
	Adj-R^2	0.029	0.024	0.022
	nobs	504	504	504
9	F	4.535	4.069	2.361
	R^2	0.021	0.026	0.012
	Adj-R^2	0.015	0.020	0.006
	nobs	504	504	504
Hi	F	2.369	4.154	0.822
	R^2	0.011	0.023	0.004
	Adj-R^2	0.006	0.017	-0.002
	nobs	504	504	504
Hi - Lo	F	15.997	4.171	18.559
	R^2	0.067	0.022	0.109
	Adj-R^2	0.061	0.017	0.103
	nobs	504	504	504
Avg	F	11.779	3.250	14.064
	R^2	0.054	0.027	0.058
	Adj-R^2	0.048	0.022	0.053
	nobs	504	504	504

Table 6: Fama-French *three* factor OLS regression on *value*-weighted portfolios sorted by elastic-net R^2. We run the Fama-French 3 factor time series regression (13) onto value-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's elastic-net R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k = \text{'Lo'}, 2, \ldots, \text{'Hi'}$ portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
EN R^2		coef	sd	t	coef	sd	t	coef	sd	t
Lo	const	0.955	0.201	(4.763)	-0.001	0.002	(-0.295)	0.963	0.200	(4.809)
	Mkt-RF	2.072	5.607	(0.370)	-0.027	0.086	(-0.318)	2.075	5.578	(0.372)
	HML	-9.290	5.471	(-1.698)	-0.085	0.138	(-0.616)	-9.203	5.489	(-1.677)
	SMB	11.577	5.524	(2.096)	0.098	0.090	(1.092)	11.480	5.516	(2.081)
2	const	0.785	0.189	(4.153)	0.002	0.004	(0.420)	0.786	0.189	(4.171)
	Mkt-RF	5.037	5.935	(0.849)	0.129	0.129	(0.994)	4.904	5.868	(0.836)
	HML	-3.615	6.971	(-0.519)	0.143	0.224	(0.638)	-3.744	6.899	(-0.543)
	SMB	4.254	5.206	(0.817)	0.227	0.134	(1.694)	4.030	5.163	(0.781)
3	const	0.501	0.187	(2.675)	0.017	0.019	(0.876)	0.487	0.183	(2.663)
	Mkt-RF	8.291	4.795	(1.729)	0.808	0.703	(1.150)	7.509	4.477	(1.677)
	HML	3.912	5.640	(0.694)	1.165	0.991	(1.175)	2.793	5.456	(0.512)
	SMB	5.327	5.165	(1.031)	0.750	0.546	(1.373)	4.661	5.145	(0.906)
4	const	0.531	0.180	(2.948)	0.023	0.042	(0.540)	0.507	0.169	(3.002)
	Mkt-RF	6.563	4.377	(1.499)	3.019	1.952	(1.547)	3.665	3.911	(0.937)
	HML	-0.399	5.381	(-0.074)	3.615	2.507	(1.442)	-4.000	5.844	(-0.685)
	SMB	0.141	4.950	(0.028)	0.630	1.022	(0.616)	-0.474	4.780	(-0.099)
5	const	0.581	0.168	(3.449)	0.015	0.060	(0.247)	0.568	0.149	(3.814)
	Mkt-RF	7.771	4.520	(1.719)	4.768	2.537	(1.879)	3.009	3.489	(0.862)
	HML	-3.689	5.101	(-0.723)	3.844	3.032	(1.268)	-7.526	4.900	(-1.536)
	SMB	0.280	4.707	(0.059)	0.355	1.360	(0.261)	-0.073	4.450	(-0.016)
6	const	0.504	0.182	(2.765)	0.041	0.079	(0.527)	0.462	0.155	(2.985)
	Mkt-RF	9.458	4.792	(1.974)	6.531	3.162	(2.065)	2.941	3.445	(0.854)
	HML	-5.354	5.077	(-1.055)	3.942	3.683	(1.070)	-9.283	4.412	(-2.104)
	SMB	-2.840	5.278	(-0.538)	-0.744	1.903	(-0.391)	-2.119	4.416	(-0.480)
7	const	0.716	0.191	(3.744)	0.074	0.097	(0.759)	0.644	0.152	(4.224)
	Mkt-RF	9.123	5.278	(1.729)	9.144	3.761	(2.431)	-0.008	3.660	(-0.002)
	HML	-10.844	5.855	(-1.852)	3.718	4.504	(0.825)	-14.524	4.792	(-3.031)
	SMB	3.499	6.268	(0.558)	-0.512	2.534	(-0.202)	4.003	5.121	(0.782)
8	const	0.628	0.210	(2.993)	0.101	0.129	(0.785)	0.529	0.154	(3.427)
	Mkt-RF	5.149	6.218	(0.828)	10.235	4.055	(2.524)	-5.071	3.820	(-1.327)
	HML	-7.949	6.411	(-1.240)	0.735	4.713	(0.156)	-8.750	4.174	(-2.096)
	SMB	1.088	7.069	(0.154)	-3.357	3.508	(-0.957)	4.430	5.400	(0.820)
9	const	0.639	0.228	(2.802)	0.094	0.162	(0.578)	0.546	0.140	(3.896)
	Mkt-RF	4.401	6.302	(0.698)	12.174	5.124	(2.376)	-7.764	3.177	(-2.444)
	HML	-9.412	6.742	(-1.396)	-2.484	6.213	(-0.400)	-6.942	3.593	(-1.932)
	SMB	0.210	5.971	(0.035)	-3.223	4.583	(-0.703)	3.437	3.700	(0.929)
Hi	const	0.631	0.247	(2.551)	0.256	0.193	(1.325)	0.376	0.124	(3.042)
	Mkt-RF	2.806	6.946	(0.404)	12.095	5.625	(2.150)	-9.252	3.145	(-2.942)
	HML	-10.997	8.361	(-1.315)	-5.184	7.981	(-0.650)	-5.758	3.624	(-1.589)
	SMB	5.037	8.083	(0.623)	2.081	6.253	(0.333)	2.922	4.027	(0.726)
Hi - Lo	const	-0.325	0.166	(-1.955)	0.257	0.193	(1.332)	-0.586	0.180	(-3.255)
	Mkt-RF	0.734	4.555	(0.161)	12.122	5.596	(2.166)	-11.327	4.597	(-2.464)
	HML	-1.707	8.217	(-0.208)	-5.099	7.966	(-0.640)	3.445	6.056	(0.569)
	SMB	-6.540	5.382	(-1.215)	1.983	6.238	(0.318)	-8.558	4.827	(-1.773)
Avg	const	0.647	0.182	(3.562)	0.062	0.073	(0.849)	0.587	0.140	(4.184)
	Mkt-RF	6.067	5.037	(1.205)	5.888	2.610	(2.256)	0.201	3.350	(0.060)
	HML	-5.764	5.096	(-1.131)	0.941	3.101	(0.303)	-6.694	3.606	(-1.856)
	SMB	2.857	5.096	(0.561)	-0.370	1.981	(-0.187)	3.230	3.937	(0.820)

EN R^2		**Actual**	**Ghost**	**Act - Gho**
Lo	F	4.573	0.768	4.516
	R^2	0.013	0.004	0.013
	Adj-R^2	0.007	-0.002	0.007
	nobs	504	504	504
2	F	1.507	1.031	1.467
	R^2	0.006	0.008	0.006
	Adj-R^2	-0.000	0.002	-0.000
	nobs	504	504	504
3	F	1.707	0.689	1.559
	R^2	0.011	0.012	0.009
	Adj-R^2	0.005	0.006	0.003
	nobs	504	504	504
4	F	0.872	0.836	0.591
	R^2	0.005	0.029	0.003
	Adj-R^2	-0.000	0.023	-0.003
	nobs	504	504	504
5	F	1.879	1.333	1.481
	R^2	0.010	0.032	0.007
	Adj-R^2	0.004	0.026	0.001
	nobs	504	504	504
6	F	2.091	1.705	1.864
	R^2	0.013	0.027	0.009
	Adj-R^2	0.007	0.021	0.003
	nobs	504	504	504
7	F	3.748	2.689	3.578
	R^2	0.019	0.033	0.017
	Adj-R^2	0.013	0.027	0.011
	nobs	504	504	504
8	F	1.326	3.009	1.660
	R^2	0.007	0.026	0.008
	Adj-R^2	0.001	0.020	0.003
	nobs	504	504	504
9	F	1.475	3.461	2.631
	R^2	0.006	0.025	0.013
	Adj-R^2	-0.000	0.019	0.007
	nobs	504	504	504
Hi	F	1.476	3.471	3.109
	R^2	0.007	0.021	0.024
	Adj-R^2	0.001	0.015	0.018
	nobs	504	504	504
Hi - Lo	F	0.527	3.491	5.041
	R^2	0.003	0.021	0.027
	Adj-R^2	-0.003	0.015	0.022
	nobs	504	504	504
Avg	F	2.408	2.899	1.987
	R^2	0.009	0.027	0.005
	Adj-R^2	0.003	0.021	-0.001
	nobs	504	504	504

Table 7: Fama-French *five* factor OLS regression on *equal*-weighted portfolios sorted by elastic-net R^2. We run the Fama-French 5 factor time series regression (14) onto equal-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's elastic-net R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
		coef	sd	t	coef	sd	t	coef	sd	t
EN R^2										
Lo	const	1.314	0.251	(5.230)	-0.002	0.002	(-0.929)	1.317	0.251	(5.243)
	Mkt-RF	24.274	6.766	(3.587)	0.051	0.090	(0.570)	24.222	6.741	(3.593)
	HML	15.472	12.946	(1.195)	-0.321	0.165	(-1.950)	15.795	12.947	(1.220)
	SMB	18.000	9.266	(1.943)	0.121	0.089	(1.368)	17.878	9.261	(1.930)
	CMA	-31.904	24.001	(-1.329)	0.509	0.189	(2.691)	-32.416	24.002	(-1.351)
	RMW	-12.451	12.985	(-0.959)	0.207	0.182	(1.136)	-12.657	12.990	(-0.974)
2	const	0.961	0.243	(3.951)	0.002	0.004	(0.623)	0.958	0.242	(3.957)
	Mkt-RF	30.814	5.896	(5.226)	0.160	0.112	(1.427)	30.655	5.862	(5.229)
	HML	18.529	11.310	(1.638)	-0.159	0.245	(-0.648)	18.700	11.287	(1.657)
	SMB	20.473	7.973	(2.568)	0.173	0.116	(1.490)	20.306	7.956	(2.552)
	CMA	-29.826	20.709	(-1.440)	0.457	0.302	(1.514)	-30.298	20.661	(-1.466)
	RMW	-9.958	11.538	(-0.863)	0.218	0.243	(0.898)	-10.176	11.543	(-0.882)
3	const	0.757	0.256	(2.956)	0.032	0.028	(1.160)	0.725	0.245	(2.966)
	Mkt-RF	28.686	6.598	(4.348)	0.942	0.730	(1.292)	27.786	6.261	(4.438)
	HML	21.818	11.865	(1.839)	2.519	1.926	(1.307)	19.315	11.461	(1.685)
	SMB	23.759	8.632	(2.752)	0.758	0.753	(1.006)	23.056	8.498	(2.713)
	CMA	-31.958	21.190	(-1.508)	-1.607	2.369	(-0.678)	-30.332	20.509	(-1.479)
	RMW	-8.682	12.104	(-0.717)	-0.200	0.986	(-0.203)	-8.404	11.925	(-0.705)
4	const	0.714	0.262	(2.724)	0.074	0.076	(0.982)	0.644	0.228	(2.819)
	Mkt-RF	28.141	6.734	(4.179)	3.825	2.341	(1.634)	24.273	5.753	(4.219)
	HML	14.848	11.026	(1.347)	7.213	5.070	(1.423)	7.408	11.024	(0.672)
	SMB	18.878	8.651	(2.182)	0.491	1.801	(0.273)	18.302	8.195	(2.233)
	CMA	-30.477	20.469	(-1.489)	-4.676	5.581	(-0.838)	-25.644	19.325	(-1.327)
	RMW	-6.391	10.936	(-0.584)	-2.000	2.721	(-0.735)	-4.344	10.247	(-0.424)
5	const	0.740	0.263	(2.820)	0.089	0.111	(0.803)	0.652	0.206	(3.171)
	Mkt-RF	23.856	7.071	(3.374)	7.070	3.620	(1.953)	16.833	5.315	(3.167)
	HML	13.011	12.142	(1.072)	7.850	6.861	(1.144)	5.259	10.453	(0.503)
	SMB	15.210	9.122	(1.667)	-0.850	2.717	(-0.313)	16.050	8.125	(1.975)
	CMA	-28.488	21.944	(-1.298)	-4.017	7.596	(-0.529)	-24.585	19.043	(-1.291)
	RMW	-7.247	10.856	(-0.668)	-3.385	3.871	(-0.874)	-3.855	9.353	(-0.412)
6	const	0.752	0.263	(2.855)	0.095	0.137	(0.692)	0.658	0.186	(3.536)
	Mkt-RF	23.437	7.152	(3.277)	10.037	4.442	(2.259)	13.435	5.186	(2.591)
	HML	9.179	12.312	(0.746)	6.863	7.997	(0.858)	2.394	10.965	(0.218)
	SMB	9.287	9.475	(0.980)	-2.754	3.465	(-0.795)	12.069	7.857	(1.536)
	CMA	-30.260	23.214	(-1.304)	-2.063	9.558	(-0.216)	-28.161	18.891	(-1.491)
	RMW	-8.056	11.620	(-0.693)	-5.142	5.044	(-1.019)	-2.926	9.714	(-0.301)
7	const	0.850	0.272	(3.126)	0.167	0.161	(1.037)	0.684	0.177	(3.867)
	Mkt-RF	19.774	7.453	(2.653)	12.575	5.110	(2.461)	7.231	4.768	(1.516)
	HML	10.298	12.387	(0.831)	5.450	9.170	(0.594)	5.093	9.711	(0.524)
	SMB	5.106	9.942	(0.514)	-5.279	4.383	(-1.204)	10.421	7.608	(1.370)
	CMA	-39.449	24.190	(-1.631)	-0.554	12.171	(-0.045)	-39.237	17.471	(-2.246)
	RMW	-17.659	11.406	(-1.548)	-8.885	5.966	(-1.489)	-8.770	8.982	(-0.976)
8	const	0.925	0.277	(3.339)	0.216	0.188	(1.152)	0.711	0.167	(4.258)
	Mkt-RF	17.714	7.680	(2.306)	15.068	5.815	(2.591)	2.723	4.396	(0.619)
	HML	8.087	13.389	(0.604)	4.130	10.416	(0.396)	3.903	9.891	(0.395)
	SMB	0.292	10.314	(0.028)	-8.504	5.533	(-1.537)	8.791	7.206	(1.220)
	CMA	-39.547	27.342	(-1.446)	-1.461	15.900	(-0.092)	-37.845	17.560	(-2.155)
	RMW	-21.507	12.616	(-1.705)	-15.028	7.726	(-1.945)	-6.620	9.551	(-0.693)
9	const	0.901	0.287	(3.140)	0.289	0.217	(1.334)	0.616	0.157	(3.923)
	Mkt-RF	12.745	8.107	(1.572)	15.317	6.665	(2.298)	-2.608	4.037	(-0.646)
	HML	4.148	14.684	(0.282)	1.160	12.144	(0.095)	2.795	9.146	(0.306)
	SMB	-2.446	10.831	(-0.226)	-11.398	7.131	(-1.598)	8.808	6.663	(1.322)
	CMA	-33.786	29.702	(-1.138)	-4.756	20.477	(-0.232)	-29.146	16.107	(-1.810)
	RMW	-23.430	14.133	(-1.658)	-21.993	9.586	(-2.294)	-1.526	9.618	(-0.159)
Hi	const	0.875	0.287	(3.046)	0.428	0.244	(1.757)	0.447	0.122	(3.657)
	Mkt-RF	7.488	8.054	(0.930)	13.887	7.165	(1.938)	-6.382	2.633	(-2.424)
	HML	6.960	15.696	(0.443)	-1.171	14.468	(-0.081)	8.152	5.902	(1.381)
	SMB	-1.843	11.058	(-0.167)	-4.001	8.827	(-0.453)	2.209	4.238	(0.521)
	CMA	-33.650	28.942	(-1.163)	-8.353	23.461	(-0.356)	-25.261	11.456	(-2.205)
	RMW	-19.552	14.027	(-1.394)	-20.637	11.137	(-1.853)	0.895	5.259	(0.170)
Hi - Lo	const	-0.440	0.157	(-2.802)	0.431	0.243	(1.770)	-0.869	0.232	(-3.751)
	Mkt-RF	-16.786	3.604	(-4.658)	13.836	7.134	(1.939)	-30.604	6.128	(-4.994)
	HML	-8.512	7.894	(-1.078)	-0.850	14.450	(-0.059)	-7.643	12.105	(-0.631)
	SMB	-19.844	6.843	(-2.900)	-4.122	8.811	(-0.468)	-15.669	7.875	(-1.990)
	CMA	-1.747	13.970	(-0.125)	-8.862	23.430	(-0.378)	7.154	19.511	(0.367)
	RMW	-7.101	8.611	(-0.825)	-20.844	11.091	(-1.879)	13.552	10.679	(1.269)
Avg	const	0.879	0.255	(3.442)	0.139	0.109	(1.276)	0.741	0.177	(4.184)
	Mkt-RF	21.693	6.878	(3.154)	7.893	3.429	(2.302)	13.817	4.458	(3.099)
	HML	12.235	12.066	(1.014)	3.353	6.428	(0.522)	8.881	8.853	(1.003)
	SMB	10.672	9.055	(1.178)	-3.124	3.146	(-0.993)	13.789	6.977	(1.976)
	CMA	-32.935	23.274	(-1.415)	-2.652	8.968	(-0.296)	-30.292	17.087	(-1.773)
	RMW	-13.493	11.441	(-1.179)	-7.685	4.191	(-1.833)	-5.838	9.073	(-0.643)

		Actual	**Ghost**	**Act - Gho**
EN R^2				
Lo	F	9.545	2.282	9.597
	R^2	0.080	0.018	0.080
	Adj-R^2	0.071	0.009	0.071
	nobs	504	504	504
2	F	16.279	1.666	16.457
	R^2	0.113	0.009	0.113
	Adj-R^2	0.104	-0.001	0.104
	nobs	504	504	504
3	F	12.092	0.543	13.595
	R^2	0.100	0.013	0.101
	Adj-R^2	0.090	0.003	0.092
	nobs	504	504	504
4	F	12.398	0.617	13.128
	R^2	0.089	0.022	0.089
	Adj-R^2	0.080	0.012	0.079
	nobs	504	504	504
5	F	8.391	0.921	9.192
	R^2	0.065	0.027	0.062
	Adj-R^2	0.055	0.017	0.052
	nobs	504	504	504
6	F	8.017	1.362	7.402
	R^2	0.057	0.031	0.054
	Adj-R^2	0.047	0.021	0.045
	nobs	504	504	504
7	F	7.631	2.023	7.567
	R^2	0.055	0.033	0.058
	Adj-R^2	0.045	0.023	0.048
	nobs	504	504	504
8	F	5.946	2.787	4.281
	R^2	0.045	0.036	0.046
	Adj-R^2	0.036	0.026	0.036
	nobs	504	504	504
9	F	3.339	3.102	1.955
	R^2	0.030	0.034	0.024
	Adj-R^2	0.020	0.024	0.015
	nobs	504	504	504
Hi	F	1.748	2.705	2.096
	R^2	0.019	0.029	0.022
	Adj-R^2	0.010	0.019	0.012
	nobs	504	504	504
Hi - Lo	F	9.990	2.718	11.143
	R^2	0.069	0.029	0.112
	Adj-R^2	0.059	0.019	0.103
	nobs	504	504	504
Avg	F	8.069	2.514	10.954
	R^2	0.061	0.032	0.069
	Adj-R^2	0.052	0.022	0.059
	nobs	504	504	504

Table 8: Fama-French *five* factor OLS regression on *value*-weighted portfolios sorted by elastic-net R^2. We run the Fama-French 5 factor time series regression (14) onto value-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's elastic-net R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten k = 'Lo', 2, ..., 'Hi' portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
		coef	sd	t	coef	sd	t	coef	sd	t
EN R^2										
Lo	const	1.051	0.205	(5.120)	-0.003	0.002	(-1.104)	1.061	0.205	(5.180)
	Mkt-RF	-2.007	5.545	(-0.362)	0.045	0.091	(0.494)	-2.079	5.514	(-0.377)
	HML	6.369	8.813	(0.723)	-0.343	0.168	(-2.041)	6.724	8.824	(0.762)
	SMB	11.222	6.448	(1.740)	0.123	0.090	(1.368)	11.098	6.434	(1.725)
	CMA	-34.585	17.382	(-1.990)	0.557	0.196	(2.839)	-35.163	17.383	(-2.023)
	RMW	-7.206	9.512	(-0.758)	0.185	0.184	(1.002)	-7.401	9.536	(-0.776)
2	const	0.845	0.197	(4.289)	-0.000	0.004	(-0.109)	0.849	0.197	(4.320)
	Mkt-RF	2.589	5.833	(0.444)	0.192	0.126	(1.527)	2.389	5.777	(0.414)
	HML	5.342	10.278	(0.520)	-0.043	0.299	(-0.144)	5.410	10.157	(0.533)
	SMB	3.613	5.891	(0.613)	0.290	0.154	(1.889)	3.323	5.835	(0.570)
	CMA	-19.499	14.823	(-1.315)	0.373	0.326	(1.143)	-19.895	14.754	(-1.349)
	RMW	-5.624	7.398	(-0.760)	0.287	0.218	(1.314)	-5.926	7.408	(-0.800)
3	const	0.553	0.192	(2.880)	0.021	0.019	(1.103)	0.536	0.188	(2.848)
	Mkt-RF	5.543	5.002	(1.108)	0.619	0.587	(1.054)	4.930	4.777	(1.032)
	HML	16.087	7.395	(2.175)	1.930	1.511	(1.277)	14.292	7.307	(1.956)
	SMB	6.673	6.251	(1.067)	0.773	0.642	(1.203)	5.992	6.216	(0.964)
	CMA	-27.938	14.425	(-1.937)	-1.717	1.775	(-0.967)	-26.425	14.300	(-1.848)
	RMW	-0.044	7.010	(-0.006)	-0.217	0.702	(-0.308)	0.163	6.999	(0.023)
4	const	0.559	0.188	(2.971)	0.033	0.043	(0.777)	0.525	0.175	(2.996)
	Mkt-RF	4.820	4.931	(0.977)	2.597	1.653	(1.571)	2.339	4.423	(0.529)
	HML	7.968	8.112	(0.982)	5.161	3.681	(1.402)	2.846	9.408	(0.302)
	SMB	1.624	5.837	(0.278)	0.519	1.197	(0.433)	1.121	5.582	(0.201)
	CMA	-19.563	14.918	(-1.311)	-3.366	3.634	(-0.926)	-16.251	15.408	(-1.055)
	RMW	1.884	7.818	(0.241)	-0.971	1.511	(-0.643)	2.849	7.739	(0.368)
5	const	0.664	0.172	(3.858)	0.026	0.062	(0.426)	0.640	0.149	(4.300)
	Mkt-RF	4.174	4.907	(0.851)	4.349	2.308	(1.884)	-0.169	3.896	(-0.043)
	HML	10.277	7.977	(1.288)	5.214	4.284	(1.217)	5.077	8.504	(0.597)
	SMB	0.120	5.418	(0.022)	0.086	1.668	(0.051)	0.039	4.950	(0.008)
	CMA	-30.948	13.827	(-2.238)	-2.872	4.363	(-0.658)	-28.094	13.333	(-2.107)
	RMW	-5.892	6.436	(-0.915)	-1.447	2.180	(-0.664)	-4.435	6.034	(-0.735)
6	const	0.605	0.184	(3.289)	0.058	0.087	(0.673)	0.547	0.150	(3.651)
	Mkt-RF	5.342	4.853	(1.101)	6.067	2.923	(2.076)	-0.710	3.377	(-0.210)
	HML	9.819	7.704	(1.275)	4.915	5.555	(0.885)	4.918	7.849	(0.627)
	SMB	-3.810	5.971	(-0.638)	-1.571	2.233	(-0.704)	-2.261	4.888	(-0.462)
	CMA	-33.107	15.091	(-2.194)	-1.630	6.389	(-0.255)	-31.480	13.977	(-2.252)
	RMW	-9.131	7.569	(-1.206)	-3.207	3.152	(-1.017)	-5.916	6.646	(-0.890)
7	const	0.881	0.199	(4.428)	0.104	0.107	(0.975)	0.779	0.152	(5.130)
	Mkt-RF	3.335	5.246	(0.636)	8.358	3.536	(2.364)	-5.016	3.663	(-1.369)
	HML	7.403	8.163	(0.907)	5.171	6.581	(0.786)	2.279	7.443	(0.306)
	SMB	-0.876	6.897	(-0.127)	-2.101	2.984	(-0.704)	1.206	5.515	(0.219)
	CMA	-37.740	15.717	(-2.401)	-2.204	7.756	(-0.284)	-35.551	12.814	(-2.774)
	RMW	-21.979	8.240	(-2.667)	-6.004	3.906	(-1.537)	-16.017	6.743	(-2.375)
8	const	0.765	0.211	(3.618)	0.151	0.134	(1.133)	0.616	0.153	(4.029)
	Mkt-RF	0.429	6.197	(0.069)	8.949	3.809	(2.350)	-8.512	3.940	(-2.160)
	HML	6.725	9.302	(0.723)	3.024	6.815	(0.444)	3.651	6.592	(0.554)
	SMB	-2.678	7.605	(-0.352)	-6.050	4.015	(-1.507)	3.341	5.595	(0.597)
	CMA	-30.191	16.697	(-1.808)	-3.348	9.727	(-0.344)	-26.866	11.877	(-2.262)
	RMW	-18.524	8.893	(-2.083)	-10.106	4.970	(-2.033)	-8.479	6.743	(-1.257)
9	const	0.772	0.229	(3.369)	0.185	0.165	(1.123)	0.587	0.139	(4.212)
	Mkt-RF	-0.651	6.277	(-0.104)	9.545	4.793	(1.991)	-10.183	3.214	(-3.169)
	HML	8.108	10.021	(0.809)	3.569	9.019	(0.396)	4.517	5.744	(0.786)
	SMB	-2.058	7.531	(-0.273)	-7.390	5.239	(-1.411)	5.344	4.322	(1.236)
	CMA	-37.486	19.307	(-1.942)	-10.763	14.142	(-0.761)	-26.710	10.803	(-2.472)
	RMW	-14.475	9.786	(-1.479)	-16.601	6.880	(-2.413)	2.152	5.534	(0.389)
Hi	const	0.740	0.249	(2.967)	0.341	0.197	(1.733)	0.400	0.128	(3.121)
	Mkt-RF	-0.703	7.037	(-0.100)	9.594	5.449	(1.761)	-10.245	3.284	(-3.119)
	HML	-1.089	11.922	(-0.091)	0.892	10.818	(0.082)	-1.990	4.995	(-0.398)
	SMB	1.260	8.571	(0.147)	-1.570	6.880	(-0.228)	2.794	4.232	(0.660)
	CMA	-19.585	21.712	(-0.902)	-11.149	17.608	(-0.633)	-8.295	8.702	(-0.953)
	RMW	-16.738	10.694	(-1.565)	-14.841	8.939	(-1.660)	-1.879	5.013	(-0.375)
Hi - Lo	const	-0.311	0.160	(-1.944)	0.344	0.196	(1.751)	-0.661	0.187	(-3.539)
	Mkt-RF	1.304	4.848	(0.269)	9.549	5.416	(1.763)	-8.166	4.578	(-1.784)
	HML	-7.458	10.240	(-0.728)	1.235	10.800	(0.114)	-8.714	8.036	(-1.084)
	SMB	-9.962	5.427	(-1.836)	-1.693	6.858	(-0.247)	-8.303	5.636	(-1.473)
	CMA	14.999	14.358	(1.045)	-11.706	17.570	(-0.666)	26.868	14.266	(1.883)
	RMW	-9.531	8.262	(-1.154)	-15.026	8.903	(-1.688)	5.522	9.407	(0.587)
Avg	const	0.744	0.187	(3.985)	0.092	0.077	(1.199)	0.654	0.142	(4.606)
	Mkt-RF	2.287	5.182	(0.441)	5.032	2.406	(2.092)	-2.726	3.577	(-0.762)
	HML	7.701	7.589	(1.015)	2.949	4.472	(0.659)	4.772	5.920	(0.806)
	SMB	1.509	6.039	(0.250)	-1.689	2.268	(-0.745)	3.200	4.658	(0.687)
	CMA	-29.064	14.597	(-1.991)	-3.612	5.920	(-0.610)	-25.473	11.462	(-2.222)
	RMW	-9.773	7.111	(-1.374)	-5.292	2.840	(-1.864)	-4.489	5.643	(-0.795)

		Actual	**Ghost**	**Act - Gho**
EN R^2				
Lo	F	3.227	2.333	3.238
	R^2	0.025	0.020	0.026
	Adj-R^2	0.016	0.010	0.016
	nobs	504	504	504
2	F	1.509	2.574	1.527
	R^2	0.010	0.012	0.010
	Adj-R^2	0.000	0.002	0.000
	nobs	504	504	504
3	F	2.053	0.436	1.883
	R^2	0.019	0.015	0.017
	Adj-R^2	0.010	0.005	0.007
	nobs	504	504	504
4	F	1.097	0.530	0.708
	R^2	0.010	0.032	0.007
	Adj-R^2	0.000	0.022	-0.003
	nobs	504	504	504
5	F	3.067	0.881	2.448
	R^2	0.022	0.033	0.019
	Adj-R^2	0.013	0.023	0.009
	nobs	504	504	504
6	F	2.060	1.203	1.873
	R^2	0.026	0.028	0.024
	Adj-R^2	0.017	0.019	0.014
	nobs	504	504	504
7	F	4.624	1.968	4.729
	R^2	0.042	0.036	0.042
	Adj-R^2	0.032	0.026	0.033
	nobs	504	504	504
8	F	1.872	2.202	2.407
	R^2	0.020	0.032	0.022
	Adj-R^2	0.010	0.023	0.012
	nobs	504	504	504
9	F	1.699	2.580	3.156
	R^2	0.019	0.036	0.028
	Adj-R^2	0.009	0.026	0.018
	nobs	504	504	504
Hi	F	1.258	2.104	2.137
	R^2	0.013	0.027	0.026
	Adj-R^2	0.003	0.017	0.016
	nobs	504	504	504
Hi - Lo	F	0.977	2.123	3.236
	R^2	0.010	0.027	0.037
	Adj-R^2	-0.000	0.017	0.027
	nobs	504	504	504
Avg	F	2.484	2.096	2.149
	R^2	0.021	0.032	0.018
	Adj-R^2	0.011	0.022	0.008
	nobs	504	504	504

Table 9: Time series regression of the cross-sectional elastic-net, FF3 and FF5 R^2 averages onto macro and financial variables. At the end of month $t-1$, we compute for each stock i the $\mathsf{R}^2{}_{i,t-1}$ coefficients under all of the elastic-net estimation method of Section 2, and also under the FFz methods of Section 5. We then compute the simple *cross-sectional average* of all the $\mathsf{R}^2{}_{i,t-1}$ across i's; that is, we compute $\bar{\mathsf{R}}^2_{t-1} := \frac{1}{N_{t-1}} \sum_{i=1}^{N_{t-1}} \mathsf{R}^2{}_{i,t-1}$, where N_{t-1} is the total number of traded stocks at month $t-1$, subject to the minimum trading days requirement outlined in Section 2. Because it is possible that $\mathsf{R}^2{}_{i,t-1}$ can take on negative values, we define the differenced quantity as $\Delta(\bar{\mathsf{R}}^2)_t := \bar{\mathsf{R}}^2_t - \bar{\mathsf{R}}^2_{t-1}$ instead of a log difference for the macro regressors. The macroeconomic regressors are: INDPRO (industrial production index), PCE (personal consumption expendistures), UNRATE (civilian unemployment rate), and PAYEMS (all employees: total nonfarm payrolls). PCE level is deflated to real terms using CPILFESL (consumer price index for all urban consumers: all items less food and energy). We denote a differenced variable by Δ; for instance, $\Delta\text{INDPRO}_t := \log(\text{INDPRO}_t) - \log(\text{INDPRO}_{t-1})$. All of the macroeconomic regressors are in log differences, and the data are from St. Louis Federal Reserve FRED Economic Data. The financial regressors are MktRF (market factor) from Kenneth French's website, and LIQ is the Pástor and Stambaugh (2003) liquidity factor. Parentheses show the Newey and West (1987) robust t-statistics with 6 lags. The sample is from December 1975 to December 2017.

	Macro regressors			**Financial regressors**			**All regressors**		
	(1) EN $\Delta(\bar{\mathsf{R}}^2)_t$	(2) FF3 $\Delta(\bar{\mathsf{R}}^2)_t$	(3) FF5 $\Delta(\bar{\mathsf{R}}^2)_t$	(4) EN $\Delta(\bar{\mathsf{R}}^2)_t$	(5) FF3 $\Delta(\bar{\mathsf{R}}^2)_t$	(6) FF5 $\Delta(\bar{\mathsf{R}}^2)_t$	(7) EN $\Delta(\bar{\mathsf{R}}^2)_t$	(8) FF3 $\Delta(\bar{\mathsf{R}}^2)_t$	(9) FF5 $\Delta(\bar{\mathsf{R}}^2)_t$
ΔINDPRO_t	0.411	0.184	0.168				0.295	0.114	0.099
	(1.480)	(1.344)	(1.239)				(1.522)	(1.233)	(1.077)
ΔPCE_t	−0.665	−0.414	−0.408				−0.473	−0.307	−0.300
	(−1.957)	(−2.576)	(−2.568)				(−2.052)	(−2.842)	(−2.808)
ΔUNRATE_t	0.067	0.031	0.032				0.074	0.034	0.035
	(1.186)	(1.096)	(1.113)				(1.475)	(1.358)	(1.389)
ΔPAYEMS_t	1.268	0.541	0.560				1.343	0.589	0.607
	(1.026)	(0.950)	(0.990)				(1.138)	(1.114)	(1.160)
$\Delta\text{INDPRO}_{t-1}$	−0.014	−0.032	−0.029				0.103	0.034	0.038
	(−0.051)	(−0.225)	(−0.209)				(0.442)	(0.289)	(0.326)
ΔPCE_{t-1}	−0.459	−0.259	−0.259				−0.286	−0.161	−0.161
	(−1.244)	(−1.542)	(−1.557)				(−0.995)	(−1.237)	(−1.241)
$\Delta\text{UNRATE}_{t-1}$	0.041	0.026	0.028				0.045	0.027	0.029
	(0.977)	(1.204)	(1.310)				(1.053)	(1.262)	(1.370)
$\Delta\text{PAYEMS}_{t-1}$	−0.868	−0.451	−0.418				−1.004	−0.533	−0.500
	(−1.127)	(−1.115)	(−1.037)				(−1.320)	(−1.322)	(−1.244)
MktRF_t				−0.178	−0.103	−0.104	−0.170	−0.098	−0.099
				(−2.046)	(−2.595)	(−2.630)	(−2.141)	(−2.705)	(−2.747)
LIQ_t				−0.013	−0.003	−0.004	−0.012	−0.001	−0.002
				(−0.308)	(−0.152)	(−0.218)	(−0.283)	(−0.032)	(−0.101)
Constant	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001
	(0.929)	(1.228)	(1.229)	(0.974)	(1.034)	(1.093)	(1.430)	(1.660)	(1.680)
Observations	503	503	503	503	503	503	503	503	503
R^2	0.030	0.038	0.037	0.086	0.109	0.111	0.105	0.132	0.134
Adjusted R^2	0.014	0.022	0.021	0.082	0.106	0.108	0.087	0.115	0.116

Table 10: Regressing the post-formation *equal*-weighted "Avg" portfolio bin returns sorted by elastic-net, FF3 and FF5 R^2 against macroeconomic shocks and financial factors. At the end of month $t-1$, we compute for each stock i the $\mathsf{R}^2{}_{i,t-1}$ coefficients under all of the elastic-net estimation method of Section 2, and also under the FFz methods of Section 5. Stocks are then sorted into decile bins according to the R^2 of a method $m \in \{\text{EN}, \text{FF3}, \text{FF5}\}$, and equal-weighted portfolios are constructed within each bin. The 'Avg' portfolio of method m is the simple average over the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' equal-weighted portfolios of their respective method. 'Act' refers to the portfolio returns of actual stocks (7), 'Gho' refers to portfolio returns of the ghost stocks (8) for $m =$ EN and (16) for $m =$ FF3, FF5, and 'Act - Gho' is the difference. The macroeconomic regressors are: INDPRO (industrial production index), PCE (personal consumption expendistures), UNRATE (civilian unemployment rate), and PAYEMS (all employees: total nonfarm payrolls). PCE level is deflated to real terms using CPILFESL (consumer price index for all urban consumers: all items less food and energy). We denote a differenced variable by Δ; for instance, $\Delta\text{INDPRO}_t := \log(\text{INDPRO}_t) - \log(\text{INDPRO}_{t-1})$. All of the macroeconomic regressors are in log differences, and the data are from St. Louis Federal Reserve FRED Economic Data. The financial regressors are MktRF (market factor) from Kenneth French's website, and LIQ is the Pástor and Stambaugh (2003) liquidity factor. Parentheses show the Newey and West (1987) robust t-statistics with 6 lags. The sample is from December 1975 to December 2017.

	EN Avg			**FF3 Avg**			**FF5 Avg**		
	(1) Act	(2) Gho	(3) Act - Gho	(4) Act	(5) Gho	(6) Act - Gho	(7) Act	(8) Gho	(9) Act - Gho
ΔINDPRO_t	0.107	0.401	−0.292	0.108	0.401	−0.290	0.108	0.401	−0.290
	(0.151)	(1.229)	(−0.556)	(0.153)	(1.229)	(−0.554)	(0.153)	(1.229)	(−0.553)
ΔPCE_t	2.432	0.324	2.104	2.434	0.324	2.105	2.434	0.324	2.105
	(1.935)	(0.827)	(2.160)	(1.936)	(0.827)	(2.160)	(1.936)	(0.827)	(2.160)
ΔUNRATE_t	−0.042	−0.016	−0.025	−0.041	−0.016	−0.025	−0.041	−0.016	−0.025
	(−0.262)	(−0.254)	(−0.235)	(−0.262)	(−0.254)	(−0.235)	(−0.262)	(−0.254)	(−0.236)
ΔPAYEMS_t	−5.240	−2.137	−3.120	−5.246	−2.137	−3.126	−5.246	−2.137	−3.126
	(−2.815)	(−2.140)	(−2.622)	(−2.818)	(−2.140)	(−2.626)	(−2.818)	(−2.140)	(−2.626)
$\Delta\text{INDPRO}_{t-1}$	0.779	−0.041	0.819	0.779	−0.041	0.818	0.779	−0.041	0.818
	(1.107)	(−0.139)	(1.592)	(1.107)	(−0.139)	(1.592)	(1.106)	(−0.139)	(1.591)
ΔPCE_{t-1}	−1.759	0.085	−1.841	−1.758	0.085	−1.839	−1.759	0.085	−1.839
	(−1.396)	(0.205)	(−1.887)	(−1.396)	(0.205)	(−1.883)	(−1.396)	(0.205)	(−1.884)
$\Delta\text{UNRATE}_{t-1}$	0.158	0.052	0.106	0.159	0.052	0.107	0.159	0.052	0.107
	(0.860)	(0.725)	(0.864)	(0.863)	(0.725)	(0.867)	(0.863)	(0.725)	(0.868)
$\Delta\text{PAYEMS}_{t-1}$	3.464	0.906	2.563	3.460	0.906	2.559	3.458	0.906	2.558
	(2.217)	(1.044)	(2.437)	(2.212)	(1.044)	(2.434)	(2.211)	(1.044)	(2.432)
MktRF_t	0.282	0.099	0.183	0.283	0.099	0.184	0.283	0.099	0.184
	(4.466)	(2.691)	(4.945)	(4.483)	(2.691)	(4.970)	(4.482)	(2.691)	(4.968)
LIQ_t	−0.099	−0.056	−0.043	−0.099	−0.056	−0.043	−0.099	−0.056	−0.043
	(−1.225)	(−1.417)	(−0.805)	(−1.231)	(−1.417)	(−0.811)	(−1.231)	(−1.417)	(−0.811)
Constant	0.007	0.001	0.006	0.007	0.001	0.006	0.007	0.001	0.006
	(1.748)	(0.585)	(2.421)	(1.747)	(0.585)	(2.419)	(1.748)	(0.585)	(2.419)
Observations	504	504	504	504	504	504	504	504	504
R^2	0.074	0.051	0.076	0.074	0.051	0.077	0.074	0.051	0.077
Adjusted R^2	0.055	0.031	0.057	0.056	0.031	0.058	0.056	0.031	0.058

Table 11: Regressing the post-formation *value*-weighted "Avg" portfolio bin returns sorted by elastic-net, FF3 and FF5 R^2 against macroeconomic shocks and financial factors. At the end of month $t-1$, we compute for each stock i the $\mathsf{R}^2{}_{i,t-1}$ coefficients under all of the elastic-net estimation method of Section 2, and also under the FFz methods of Section 5. Stocks are then sorted into decile bins according to the R^2 of a method $m \in \{\text{EN}, \text{FF3}, \text{FF5}\}$, and value-weighted portfolios are constructed within each bin. The 'Avg' portfolio of method m is the simple average over the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' value-weighted portfolios of their respective method. 'Act' refers to the portfolio returns of actual stocks (7), 'Gho' refers to portfolio returns of the ghost stocks (8) for $m =$ EN and (16) for $m =$ FF3, FF5, and 'Act - Gho' is the difference. The macroeconomic regressors are: INDPRO (industrial production index), PCE (personal consumption expendistures), UNRATE (civilian unemployment rate), and PAYEMS (all employees: total nonfarm payrolls). PCE level is deflated to real terms using CPILFESL (consumer price index for all urban consumers: all items less food and energy). We denote a differenced variable by Δ; for instance, $\Delta\text{INDPRO}_t := \log(\text{INDPRO}_t) - \log(\text{INDPRO}_{t-1})$. All of the macroeconomic regressors are in log differences, and the data are from St. Louis Federal Reserve FRED Economic Data. The financial regressors are MktRF (market factor) from Kenneth French's website, and LIQ is the Pástor and Stambaugh (2003) liquidity factor. Parentheses show the Newey and West (1987) robust t-statistics with 6 lags. The sample is from December 1975 to December 2017.

	EN Avg			**FF3 Avg**			**FF5 Avg**		
	(1) Act	(2) Gho	(3) Act - Gho	(4) Act	(5) Gho	(6) Act - Gho	(7) Act	(8) Gho	(9) Act - Gho
ΔINDPRO_t	0.175	0.230	−0.058	0.124	0.292	−0.171	0.042	0.256	−0.216
	(0.323)	(0.968)	(−0.136)	(0.241)	(1.199)	(−0.432)	(0.081)	(1.096)	(−0.530)
ΔPCE_t	2.137	0.315	1.823	1.917	0.282	1.635	2.004	0.286	1.717
	(2.277)	(1.238)	(2.350)	(1.955)	(1.181)	(2.029)	(2.066)	(1.293)	(2.111)
ΔUNRATE_t	0.054	−0.015	0.070	0.033	−0.009	0.042	0.017	−0.009	0.026
	(0.490)	(−0.331)	(0.785)	(0.312)	(−0.208)	(0.527)	(0.157)	(−0.219)	(0.333)
ΔPAYEMS_t	−3.283	−1.561	−1.719	−3.545	−1.562	−1.977	−3.604	−1.498	−2.099
	(−2.307)	(−2.321)	(−1.525)	(−2.589)	(−2.344)	(−1.904)	(−2.670)	(−2.265)	(−2.027)
$\Delta\text{INDPRO}_{t-1}$	0.590	0.024	0.566	0.625	−0.027	0.650	0.726	−0.008	0.733
	(1.032)	(0.110)	(1.242)	(1.175)	(−0.123)	(1.596)	(1.358)	(−0.038)	(1.751)
ΔPCE_{t-1}	−1.779	−0.035	−1.744	−1.504	0.024	−1.525	−1.660	0.006	−1.664
	(−1.933)	(−0.130)	(−2.335)	(−1.570)	(0.095)	(−1.937)	(−1.750)	(0.026)	(−2.085)
$\Delta\text{UNRATE}_{t-1}$	−0.051	0.028	−0.079	−0.004	0.022	−0.026	0.015	0.022	−0.007
	(−0.403)	(0.561)	(−0.824)	(−0.033)	(0.472)	(−0.303)	(0.130)	(0.490)	(−0.082)
$\Delta\text{PAYEMS}_{t-1}$	1.907	0.717	1.194	2.089	0.601	1.495	1.926	0.548	1.384
	(1.546)	(1.233)	(1.068)	(1.717)	(1.028)	(1.431)	(1.641)	(0.941)	(1.367)
MktRF_t	0.074	0.066	0.009	0.145	0.071	0.075	0.146	0.067	0.080
	(1.604)	(2.479)	(0.285)	(3.062)	(2.477)	(2.612)	(3.090)	(2.425)	(2.677)
LIQ_t	−0.021	−0.034	0.013	−0.033	−0.031	−0.003	−0.032	−0.032	0.0003
	(−0.397)	(−1.262)	(0.334)	(−0.622)	(−1.112)	(−0.069)	(−0.583)	(−1.192)	(0.008)
Constant	0.006	0.001	0.005	0.005	0.001	0.004	0.006	0.001	0.005
	(2.025)	(0.521)	(2.711)	(1.837)	(0.609)	(2.380)	(2.029)	(0.658)	(2.550)
Observations	504	504	504	504	504	504	504	504	504
R^2	0.043	0.049	0.043	0.057	0.056	0.049	0.057	0.052	0.052
Adjusted R^2	0.024	0.030	0.023	0.038	0.036	0.030	0.038	0.033	0.033

Table 12: Regressing the post-formation *equal*-weighted "Hi - Lo" portfolio bin returns sorted by elastic-net, FF3 and FF5 R^2 against macroeconomic shocks and financial factors. At the end of month $t-1$, we compute for each stock i the $\mathsf{R}^2{}_{i,t-1}$ coefficients under all of the elastic-net estimation method of Section 2, and also under the FFz methods of Section 5. Stocks are then sorted into decile bins according to the R^2 of a method $m \in \{\text{EN}, \text{FF3}, \text{FF5}\}$, and equal-weighted portfolios are constructed within each bin. 'Lo' refers to the lowest decile portfolio, 'Hi' to the highest decile portfolio, and 'Hi - Lo' refers to the difference. 'Act' refers to the portfolio returns of actual stocks (7), 'Gho' refers to portfolio returns of the ghost stocks (8) for $m = \text{EN}$ and (16) for $m = \text{FF3}, \text{FF5}$, and 'Act - Gho' is the difference. The macroeconomic regressors are: INDPRO (industrial production index), PCE (personal consumption expendistures), UNRATE (civilian unemployment rate), and PAYEMS (all employees: total nonfarm payrolls). PCE level is deflated to real terms using CPILFESL (consumer price index for all urban consumers: all items less food and energy). We denote a differenced variable by Δ; for instance, $\Delta\text{INDPRO}_t := \log(\text{INDPRO}_t) - \log(\text{INDPRO}_{t-1})$. All of the macroeconomic regressors are in log differences, and the data are from St. Louis Federal Reserve FRED Economic Data. The financial regressors are MktRF (market factor) from Kenneth French's website, and LIQ is the Pástor and Stambaugh (2003) liquidity factor. Parentheses show the Newey and West (1987) robust t-statistics with 6 lags. The sample is from December 1975 to December 2017.

	EN Hi - Lo			**FF3 Hi - Lo**			**FF5 Hi - Lo**		
	(1) Act	(2) Gho	(3) Act - Gho	(4) Act	(5) Gho	(6) Act - Gho	(7) Act	(8) Gho	(9) Act - Gho
ΔINDPRO_t	0.720	−0.814	1.516	1.580	−1.098	2.675	1.305	−1.013	2.316
	(0.994)	(−0.455)	(0.925)	(1.214)	(−0.674)	(1.370)	(1.017)	(−0.616)	(1.213)
ΔPCE_t	0.829	5.523	−4.706	0.514	5.277	−4.756	1.057	5.464	−4.404
	(1.242)	(2.582)	(−2.630)	(0.323)	(2.572)	(−1.902)	(0.675)	(2.683)	(−1.820)
ΔUNRATE_t	0.344	−0.050	0.399	0.106	−0.129	0.236	−0.041	−0.154	0.114
	(1.600)	(−0.119)	(0.956)	(0.158)	(−0.445)	(0.284)	(−0.063)	(−0.526)	(0.144)
ΔPAYEMS_t	−2.617	−10.346	7.844	−6.434	−12.705	6.348	−7.242	−13.051	5.907
	(−0.332)	(−1.079)	(0.900)	(−0.731)	(−1.528)	(0.587)	(−0.824)	(−1.549)	(0.569)
$\Delta\text{INDPRO}_{t-1}$	−0.543	1.532	−2.057	−1.702	1.652	−3.357	−1.387	1.586	−2.974
	(−0.756)	(0.844)	(−1.265)	(−1.370)	(0.995)	(−1.790)	(−1.145)	(0.951)	(−1.617)
ΔPCE_{t-1}	−1.035	−4.975	3.951	−1.247	−4.845	3.594	−1.700	−5.043	3.339
	(−1.506)	(−2.244)	(2.172)	(−0.766)	(−2.278)	(1.411)	(−1.067)	(−2.387)	(1.355)
$\Delta\text{UNRATE}_{t-1}$	−0.375	0.116	−0.493	−0.220	0.162	−0.384	−0.061	0.190	−0.252
	(−1.669)	(0.272)	(−1.165)	(−0.330)	(0.565)	(−0.462)	(−0.095)	(0.658)	(−0.317)
$\Delta\text{PAYEMS}_{t-1}$	3.362	8.682	−5.448	8.120	11.589	−3.520	8.901	11.956	−3.130
	(0.430)	(0.936)	(−0.593)	(0.934)	(1.436)	(−0.329)	(1.021)	(1.465)	(−0.302)
MktRF_t	−0.213	0.194	−0.407	−0.395	0.135	−0.530	−0.391	0.141	−0.532
	(−6.710)	(3.029)	(−6.544)	(−7.591)	(2.575)	(−8.593)	(−7.705)	(2.543)	(−8.969)
LIQ_t	−0.021	−0.113	0.092	0.155	−0.095	0.250	0.141	−0.096	0.237
	(−0.343)	(−1.364)	(1.286)	(1.844)	(−1.252)	(3.102)	(1.665)	(−1.238)	(3.040)
Constant	−0.005	0.003	−0.008	−0.003	0.001	−0.004	−0.003	0.001	−0.005
	(−2.232)	(0.818)	(−1.931)	(−0.948)	(0.439)	(−1.128)	(−1.030)	(0.416)	(−1.198)
Observations	504	504	504	504	504	504	504	504	504
R^2	0.076	0.055	0.149	0.127	0.048	0.208	0.128	0.050	0.215
Adjusted R^2	0.057	0.036	0.132	0.109	0.029	0.192	0.110	0.031	0.199

Table 13: Regressing the post-formation *value*-weighted "Hi - Lo" portfolio bin returns sorted by elastic-net, FF3 and FF5 R^2 against macroeconomic shocks and financial factors. At the end of month $t-1$, we compute for each stock i the $\mathsf{R}^2{}_{i,t-1}$ coefficients under all of the elastic-net estimation method of Section 2, and also under the FFz methods of Section 5. Stocks are then sorted into decile bins according to the R^2 of a method $m \in \{\text{EN}, \text{FF3}, \text{FF5}\}$, and value-weighted portfolios are constructed within each bin. 'Lo' refers to the lowest decile portfolio, 'Hi' to the highest decile portfolio, and 'Hi - Lo' refers to the difference. 'Act' refers to the portfolio returns of actual stocks (7), 'Gho' refers to portfolio returns of the ghost stocks (8) for $m = \text{EN}$ and (16) for $m = \text{FF3}, \text{FF5}$, and 'Act - Gho' is the difference. The macroeconomic regressors are: INDPRO (industrial production index), PCE (personal consumption expendistures), UNRATE (civilian unemployment rate), and PAYEMS (all employees: total nonfarm payrolls). PCE level is deflated to real terms using CPILFESL (consumer price index for all urban consumers: all items less food and energy). We denote a differenced variable by Δ; for instance, $\Delta\text{INDPRO}_t := \log(\text{INDPRO}_t) - \log(\text{INDPRO}_{t-1})$. All of the macroeconomic regressors are in log differences, and the data are from St. Louis Federal Reserve FRED Economic Data. The financial regressors are MktRF (market factor) from Kenneth French's website, and LIQ is the Pástor and Stambaugh (2003) liquidity factor. Parentheses show the Newey and West (1987) robust t-statistics with 6 lags. The sample is from December 1975 to December 2017.

	EN Hi - Lo			**FF3 Hi - Lo**			**FF5 Hi - Lo**		
	(1) Act	(2) Gho	(3) Act - Gho	(4) Act	(5) Gho	(6) Act - Gho	(7) Act	(8) Gho	(9) Act - Gho
ΔINDPRO_t	−2.700	−0.554	−2.169	−0.277	−0.953	0.678	−0.431	−0.732	0.306
	(−2.880)	(−0.434)	(−2.338)	(−0.243)	(−0.937)	(0.479)	(−0.359)	(−0.737)	(0.217)
ΔPCE_t	3.038	3.964	−0.894	0.876	3.661	−2.783	1.748	3.712	−1.968
	(2.306)	(2.598)	(−0.943)	(0.789)	(3.044)	(−1.902)	(1.684)	(3.210)	(−1.470)
ΔUNRATE_t	0.072	−0.124	0.196	0.166	−0.096	0.264	0.054	−0.129	0.185
	(0.261)	(−0.352)	(0.794)	(0.449)	(−0.421)	(0.577)	(0.147)	(−0.589)	(0.435)
ΔPAYEMS_t	2.846	−7.535	10.265	0.783	−5.860	6.611	−1.424	−6.194	4.801
	(0.468)	(−1.010)	(2.265)	(0.112)	(−0.914)	(1.030)	(−0.200)	(−0.975)	(0.802)
$\Delta\text{INDPRO}_{t-1}$	3.116	1.105	2.035	0.648	1.375	−0.727	0.995	1.267	−0.275
	(3.182)	(0.850)	(2.116)	(0.570)	(1.319)	(−0.527)	(0.819)	(1.249)	(−0.196)
ΔPCE_{t-1}	−3.128	−3.431	0.276	−1.025	−3.250	2.229	−1.920	−3.330	1.418
	(−2.426)	(−2.176)	(0.285)	(−0.943)	(−2.602)	(1.484)	(−1.905)	(−2.770)	(1.032)
$\Delta\text{UNRATE}_{t-1}$	−0.036	0.149	−0.186	−0.195	0.111	−0.309	−0.097	0.138	−0.238
	(−0.132)	(0.419)	(−0.749)	(−0.531)	(0.482)	(−0.677)	(−0.269)	(0.628)	(−0.566)
$\Delta\text{PAYEMS}_{t-1}$	−1.798	6.162	−7.847	−0.239	4.845	−5.059	1.736	4.935	−3.232
	(−0.309)	(0.864)	(−1.806)	(−0.035)	(0.798)	(−0.790)	(0.254)	(0.821)	(−0.554)
MktRF_t	−0.004	0.141	−0.145	−0.195	0.111	−0.305	−0.220	0.113	−0.333
	(−0.124)	(2.827)	(−3.273)	(−4.368)	(2.641)	(−5.657)	(−4.803)	(2.600)	(−6.332)
LIQ_t	−0.103	−0.075	−0.029	0.019	−0.053	0.071	0.020	−0.053	0.073
	(−2.167)	(−1.187)	(−0.463)	(0.299)	(−0.947)	(1.148)	(0.328)	(−0.958)	(1.197)
Constant	−0.004	0.002	−0.006	−0.001	0.001	−0.002	−0.0002	0.001	−0.001
	(−1.948)	(0.772)	(−2.068)	(−0.200)	(0.362)	(−0.516)	(−0.078)	(0.439)	(−0.493)
Observations	504	504	504	504	504	504	504	504	504
R^2	0.041	0.048	0.048	0.049	0.046	0.115	0.077	0.051	0.139
Adjusted R^2	0.022	0.029	0.029	0.029	0.027	0.097	0.059	0.032	0.121

Table 14: Unconditional correlations between elastic-net, FF3 and FF5 R^2's. We stack all the R^2's from the elastic-net, FF3 and FF5 regressions column-wise, so that each row has the form $[\mathsf{R}^2{}_{i,t}, \mathsf{R}^2{}_{\text{FF3},i,t}, \mathsf{R}^2{}_{\text{FF5},i,t}]$ in that we discard all stock i and time t information. We compute the correlation of the resulting 3-column, many rows matrix. The sample period is from December 1975 to December 2017.

	EN R^2	FF3 R^2	FF5 R^2
EN R^2	1.000		
FF3 R^2	0.811	1.000	
FF5 R^2	0.815	0.995	1.000

Table 15: Univariate equal- and value- weighted portfolio sort by Fama-French three factor model R^2. *Contrast these results against Table 3.* We form equal- and value-weighted portfolios decile portfolios every month by regressing each stock's daily return over the past year onto the daily Fama-French 3 factors using OLS. Stocks are sorted into deciles based on their FF3 R^2 from the lowest (quantile 1, labelled "Lo") to highest (quantile 10, labelled "Hi"). The column labelled "Hi - Lo" is the monthly return difference between the "Hi" bin and the "Lo" bin. The column labelled "Actual" reports the portfolio of actual stocks (7). The column labelled "Ghost" reports the portfolio of returns that are constructed out of the estimated FF3 beta coefficients according to (16). The column labelled "Act - Gho" reports the portfolio of a long position in the actual stocks, and a short position in the corresponding ghost portfolios of the actual stocks. The value weight of the ghost portfolio of stock i is the same value weight as the actual stock i. For "Actual" and "Ghost", the columns labelled "mean" and "sd" are excess returns measured in monthly percentage terms (e.g. 1.0 means 1%). For "Act - Gho", the columns labelled "mean" and "sd" are total simple returns, and not excess returns, measured in monthly percentage terms. Robust Newey and West (1987) t-statistics are reported in column "t" in parentheses. The sample period is from December 1975 to December 2017.

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
FF3 R^2									
Lo	1.207	5.021	(3.772)	0.026	1.019	(1.914)	1.183	4.953	(3.796)
2	1.092	5.017	(3.494)	0.032	1.327	(1.499)	1.063	4.892	(3.559)
3	0.972	5.095	(3.043)	0.066	1.956	(1.570)	0.905	4.854	(3.092)
4	0.952	5.199	(2.978)	0.126	2.761	(1.383)	0.828	4.772	(3.106)
5	0.941	5.225	(3.142)	0.169	3.244	(1.323)	0.775	4.638	(3.345)
6	0.876	5.183	(3.022)	0.190	3.494	(1.287)	0.687	4.466	(3.297)
7	0.888	5.163	(3.226)	0.214	3.713	(1.362)	0.677	4.313	(3.645)
8	0.894	5.204	(3.449)	0.253	3.942	(1.501)	0.641	4.225	(3.838)
9	0.885	5.242	(3.400)	0.262	4.225	(1.379)	0.624	4.083	(4.175)
Hi	0.810	5.315	(3.093)	0.277	4.663	(1.270)	0.533	3.706	(4.299)
Hi - Lo	-0.397	4.893	(-1.781)	0.251	4.593	(1.203)	-0.651	5.046	(-2.123)
Avg	0.952	5.024	(3.396)	0.161	3.170	(1.437)	0.792	4.257	(3.928)

(a) Equal-weighted

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
FF3 R^2									
Lo	0.610	4.149	(2.795)	0.030	1.415	(1.588)	0.588	4.075	(2.788)
2	0.632	4.500	(2.619)	0.023	1.811	(0.621)	0.611	4.357	(2.680)
3	0.769	4.397	(3.348)	0.016	2.093	(0.366)	0.760	4.136	(3.712)
4	0.651	4.338	(3.156)	0.085	2.325	(1.273)	0.573	4.064	(3.234)
5	0.718	4.357	(3.511)	0.142	2.665	(1.631)	0.579	3.993	(3.553)
6	0.617	4.343	(3.014)	0.148	2.772	(1.633)	0.473	3.948	(2.899)
7	0.701	4.327	(3.525)	0.119	2.897	(1.161)	0.586	3.859	(3.917)
8	0.668	4.346	(3.470)	0.159	3.100	(1.393)	0.511	3.813	(3.755)
9	0.703	4.514	(3.452)	0.206	3.478	(1.454)	0.501	3.790	(3.685)
Hi	0.594	4.688	(2.657)	0.206	4.039	(1.171)	0.389	3.494	(3.212)
Hi - Lo	-0.015	4.287	(-0.076)	0.176	3.989	(1.046)	-0.199	4.289	(-0.882)
Avg	0.666	4.192	(3.427)	0.113	2.603	(1.405)	0.557	3.660	(3.895)

(b) Value-weighted

Table 16: Univariate equal- and value- weighted portfolio sort by Fama-French five factor model R^2. *Contrast these results against Table 3.* We form equal- and value-weighted portfolios decile portfolios every month by regressing each stock's daily return over the past year onto the daily Fama-French 5 factors using OLS. Stocks are sorted into deciles based on their FF5 R^2 from the lowest (quantile 1, labelled "Lo") to highest (quantile 10, labelled "Hi"). The column labelled "Hi - Lo" is the monthly return difference between the "Hi" bin and the "Lo" bin. The column labelled "Actual" reports the portfolio of actual stocks (7). The column labelled "Ghost" reports the portfolio of returns that are constructed out of the estimated FF3 beta coefficients according to (16). The column labelled "Act - Gho" reports the portfolio of a long position in the actual stocks, and a short position in the corresponding ghost portfolios of the actual stocks. The value weight of the ghost portfolio of stock i is the same value weight as the actual stock i. For "Actual" and "Ghost", the columns labelled "mean" and "sd" are excess returns measured in monthly percentage terms (e.g. 1.0 means 1%). For "Act - Gho", the columns labelled "mean" and "sd" are total simple returns, and not excess returns, measured in monthly percentage terms. Robust Newey and West (1987) t-statistics are reported in column "t" in parentheses. The sample period is from December 1975 to December 2017.

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
FF5 R^2									
Lo	1.171	4.961	(3.776)	0.021	0.980	(1.715)	1.151	4.894	(3.810)
2	1.114	5.052	(3.496)	0.029	1.307	(1.490)	1.088	4.932	(3.565)
3	1.001	5.121	(3.124)	0.071	1.956	(1.670)	0.928	4.883	(3.159)
4	0.958	5.193	(3.002)	0.128	2.774	(1.387)	0.833	4.768	(3.128)
5	0.949	5.199	(3.159)	0.165	3.228	(1.308)	0.787	4.614	(3.369)
6	0.877	5.186	(3.032)	0.197	3.490	(1.343)	0.682	4.472	(3.259)
7	0.882	5.173	(3.217)	0.216	3.723	(1.367)	0.667	4.318	(3.618)
8	0.880	5.202	(3.395)	0.243	3.952	(1.437)	0.637	4.210	(3.875)
9	0.885	5.223	(3.410)	0.269	4.205	(1.436)	0.616	4.073	(4.085)
Hi	0.797	5.318	(3.035)	0.273	4.683	(1.232)	0.523	3.682	(4.327)
Hi - Lo	-0.374	4.830	(-1.744)	0.252	4.613	(1.184)	-0.628	4.979	(-2.108)
Avg	0.951	5.024	(3.395)	0.161	3.170	(1.437)	0.791	4.257	(3.927)

(a) Equal-weighted

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
FF5 R^2									
Lo	0.546	4.093	(2.560)	0.022	1.089	(1.800)	0.532	4.040	(2.556)
2	0.605	4.424	(2.596)	0.023	1.727	(0.641)	0.589	4.291	(2.694)
3	0.894	4.491	(3.973)	0.030	2.037	(0.676)	0.869	4.257	(4.276)
4	0.709	4.324	(3.386)	0.098	2.293	(1.493)	0.619	4.081	(3.361)
5	0.838	4.372	(4.023)	0.133	2.624	(1.545)	0.709	4.012	(4.209)
6	0.650	4.370	(3.175)	0.128	2.767	(1.425)	0.523	3.972	(3.214)
7	0.739	4.346	(3.826)	0.168	2.944	(1.664)	0.574	3.865	(3.873)
8	0.692	4.337	(3.699)	0.167	3.056	(1.526)	0.527	3.829	(3.835)
9	0.676	4.466	(3.314)	0.196	3.388	(1.449)	0.484	3.799	(3.454)
Hi	0.595	4.646	(2.677)	0.203	4.009	(1.162)	0.392	3.474	(3.277)
Hi - Lo	0.049	4.161	(0.255)	0.181	3.943	(1.081)	-0.140	4.218	(-0.636)
Avg	0.694	4.206	(3.591)	0.117	2.583	(1.478)	0.582	3.687	(4.043)

(b) Value-weighted

Table 17: Fama-French *five* factor OLS regression on *equal*-weighted portfolios sorted by FF5 R^2. *Contrast these results with Table 7.* We run the Fama-French 5 factor time series regression (14) onto equal-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's FF5 R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
		coef	sd	t	coef	sd	t	coef	sd	t
FF5 R^2										
Lo	const	0.893	0.243	(3.671)	0.022	0.018	(1.202)	0.875	0.232	(3.778)
	Mkt-RF	40.477	6.407	(6.318)	1.503	0.461	(3.258)	38.950	6.058	(6.429)
	HML	21.068	10.225	(2.060)	0.402	1.042	(0.386)	20.372	9.663	(2.108)
	SMB	31.652	8.610	(3.676)	0.784	0.512	(1.531)	30.697	8.309	(3.694)
	CMA	-29.613	18.729	(-1.581)	-0.259	1.525	(-0.170)	-28.877	17.707	(-1.631)
	RMW	-1.768	11.782	(-0.150)	-0.881	0.736	(-1.197)	-1.083	11.305	(-0.096)
2	const	0.857	0.270	(3.169)	0.035	0.034	(1.032)	0.825	0.247	(3.335)
	Mkt-RF	33.308	6.799	(4.899)	2.560	0.907	(2.821)	30.717	6.181	(4.969)
	HML	16.744	11.308	(1.481)	1.348	1.675	(0.805)	15.377	10.512	(1.463)
	SMB	25.815	9.405	(2.745)	0.993	0.958	(1.036)	24.792	8.937	(2.774)
	CMA	-26.447	20.986	(-1.260)	-1.462	2.552	(-0.573)	-24.985	19.294	(-1.295)
	RMW	-9.374	11.903	(-0.788)	-0.933	1.410	(-0.662)	-8.448	11.049	(-0.765)
3	const	0.749	0.275	(2.721)	0.064	0.067	(0.963)	0.685	0.232	(2.953)
	Mkt-RF	33.359	7.552	(4.418)	4.807	1.959	(2.454)	28.675	6.232	(4.601)
	HML	15.289	12.908	(1.184)	2.954	3.724	(0.793)	12.467	11.080	(1.125)
	SMB	19.892	9.651	(2.061)	0.118	1.888	(0.063)	19.698	8.698	(2.265)
	CMA	-32.454	22.215	(-1.461)	-2.109	5.153	(-0.409)	-30.399	19.052	(-1.596)
	RMW	-8.478	11.604	(-0.731)	-2.553	2.564	(-0.996)	-6.007	10.293	(-0.584)
4	const	0.806	0.286	(2.818)	0.104	0.114	(0.909)	0.706	0.214	(3.307)
	Mkt-RF	27.351	7.740	(3.534)	8.488	3.475	(2.443)	18.938	5.567	(3.402)
	HML	11.439	13.259	(0.863)	6.404	6.856	(0.934)	5.314	10.703	(0.496)
	SMB	14.662	10.047	(1.459)	-0.975	2.916	(-0.335)	15.776	8.518	(1.852)
	CMA	-28.485	24.951	(-1.142)	-1.854	8.573	(-0.216)	-26.874	20.184	(-1.331)
	RMW	-12.344	12.434	(-0.993)	-5.067	3.955	(-1.281)	-7.075	10.368	(-0.682)
5	const	0.878	0.276	(3.175)	0.142	0.136	(1.046)	0.739	0.194	(3.810)
	Mkt-RF	23.210	7.641	(3.038)	10.223	4.271	(2.394)	13.006	5.122	(2.539)
	HML	1.832	13.875	(0.132)	4.089	7.914	(0.517)	-2.247	10.615	(-0.212)
	SMB	6.852	10.243	(0.669)	-3.452	3.529	(-0.978)	10.208	8.231	(1.240)
	CMA	-26.104	25.476	(-1.025)	-1.056	9.909	(-0.107)	-25.222	19.483	(-1.295)
	RMW	-21.359	12.738	(-1.677)	-7.536	4.844	(-1.556)	-14.003	10.414	(-1.345)
6	const	0.875	0.286	(3.059)	0.178	0.153	(1.167)	0.699	0.192	(3.631)
	Mkt-RF	18.286	7.579	(2.413)	11.706	4.907	(2.385)	6.541	4.590	(1.425)
	HML	9.032	13.827	(0.653)	5.181	9.089	(0.570)	3.809	10.630	(0.358)
	SMB	1.267	10.660	(0.119)	-5.060	4.119	(-1.228)	6.362	8.333	(0.763)
	CMA	-35.769	26.757	(-1.337)	-2.941	11.883	(-0.247)	-32.978	19.970	(-1.651)
	RMW	-15.245	13.162	(-1.158)	-9.935	5.553	(-1.789)	-5.332	10.667	(-0.500)
7	const	0.885	0.262	(3.383)	0.186	0.164	(1.136)	0.699	0.165	(4.227)
	Mkt-RF	14.086	7.381	(1.908)	13.289	5.309	(2.503)	0.772	4.146	(0.186)
	HML	8.435	12.989	(0.649)	3.188	9.262	(0.344)	5.297	9.978	(0.531)
	SMB	0.434	10.075	(0.043)	-6.470	4.635	(-1.396)	6.858	7.280	(0.942)
	CMA	-33.698	26.364	(-1.278)	-0.352	13.213	(-0.027)	-33.466	18.857	(-1.775)
	RMW	-18.897	12.015	(-1.573)	-12.162	6.369	(-1.910)	-6.753	9.201	(-0.734)
8	const	0.952	0.267	(3.563)	0.229	0.174	(1.318)	0.726	0.163	(4.443)
	Mkt-RF	10.658	7.797	(1.367)	13.170	5.556	(2.371)	-2.523	4.153	(-0.608)
	HML	13.997	13.658	(1.025)	3.932	9.861	(0.399)	10.062	9.092	(1.107)
	SMB	-4.130	10.222	(-0.404)	-7.040	5.359	(-1.314)	2.845	6.742	(0.422)
	CMA	-40.372	27.105	(-1.489)	-3.195	14.764	(-0.216)	-37.198	18.108	(-2.054)
	RMW	-20.710	12.150	(-1.705)	-15.837	6.952	(-2.278)	-4.945	8.810	(-0.561)
9	const	0.968	0.277	(3.496)	0.269	0.194	(1.388)	0.699	0.155	(4.522)
	Mkt-RF	7.445	7.991	(0.932)	13.050	6.024	(2.166)	-5.612	3.842	(-1.461)
	HML	9.587	13.401	(0.715)	1.507	10.693	(0.141)	7.926	8.603	(0.921)
	SMB	-4.643	9.902	(-0.469)	-7.239	6.270	(-1.154)	2.647	5.855	(0.452)
	CMA	-37.900	27.763	(-1.365)	-4.180	17.474	(-0.239)	-33.550	16.413	(-2.044)
	RMW	-19.544	12.340	(-1.584)	-17.652	7.926	(-2.227)	-1.876	7.889	(-0.238)
Hi	const	0.950	0.283	(3.360)	0.309	0.224	(1.381)	0.640	0.130	(4.908)
	Mkt-RF	-0.351	7.804	(-0.045)	11.254	6.599	(1.706)	-11.615	3.343	(-3.474)
	HML	6.910	14.720	(0.469)	1.210	13.016	(0.093)	5.800	5.899	(0.983)
	SMB	-6.745	10.250	(-0.658)	-7.118	8.076	(-0.881)	0.346	4.484	(0.077)
	CMA	-39.663	27.568	(-1.439)	-11.127	22.098	(-0.504)	-28.741	11.052	(-2.600)
	RMW	-19.143	13.336	(-1.435)	-18.921	10.334	(-1.831)	-0.328	5.743	(-0.057)
Hi - Lo	const	0.057	0.213	(0.266)	0.287	0.212	(1.353)	-0.235	0.238	(-0.986)
	Mkt-RF	-40.828	6.780	(-6.022)	9.751	6.301	(1.548)	-50.564	6.890	(-7.339)
	HML	-14.158	11.119	(-1.273)	0.808	12.199	(0.066)	-14.573	10.033	(-1.453)
	SMB	-38.397	8.640	(-4.444)	-7.903	7.699	(-1.026)	-30.351	8.159	(-3.720)
	CMA	-10.050	16.259	(-0.618)	-10.868	20.992	(-0.518)	0.136	15.730	(0.009)
	RMW	-17.375	12.399	(-1.401)	-18.039	9.883	(-1.825)	0.755	11.139	(0.068)
Avg	const	0.881	0.259	(3.401)	0.154	0.122	(1.257)	0.729	0.170	(4.293)
	Mkt-RF	20.783	7.034	(2.954)	9.005	3.835	(2.348)	11.785	4.236	(2.782)
	HML	11.433	12.362	(0.925)	3.021	7.076	(0.427)	8.418	8.605	(0.978)
	SMB	8.506	9.260	(0.918)	-3.546	3.643	(-0.973)	12.023	6.793	(1.770)
	CMA	-33.051	24.049	(-1.374)	-2.853	10.269	(-0.278)	-30.229	16.704	(-1.810)
	RMW	-14.686	11.493	(-1.278)	-9.148	4.767	(-1.919)	-5.585	8.564	(-0.652)

		Actual	**Ghost**	**Act - Gho**
FF5 R^2				
Lo	F	22.748	6.584	23.607
	R^2	0.165	0.058	0.168
	Adj-R^2	0.157	0.048	0.159
	nobs	504	504	504
2	F	14.690	3.655	15.474
	R^2	0.110	0.038	0.113
	Adj-R^2	0.101	0.028	0.104
	nobs	504	504	504
3	F	11.213	2.382	13.662
	R^2	0.102	0.027	0.110
	Adj-R^2	0.093	0.018	0.101
	nobs	504	504	504
4	F	9.002	1.899	10.984
	R^2	0.068	0.032	0.071
	Adj-R^2	0.059	0.022	0.062
	nobs	504	504	504
5	F	8.423	2.350	9.389
	R^2	0.063	0.033	0.065
	Adj-R^2	0.053	0.023	0.055
	nobs	504	504	504
6	F	4.340	2.376	3.172
	R^2	0.042	0.036	0.036
	Adj-R^2	0.032	0.026	0.026
	nobs	504	504	504
7	F	4.438	2.794	2.812
	R^2	0.033	0.037	0.027
	Adj-R^2	0.024	0.028	0.017
	nobs	504	504	504
8	F	2.938	2.991	1.459
	R^2	0.027	0.036	0.021
	Adj-R^2	0.017	0.026	0.011
	nobs	504	504	504
9	F	1.985	2.909	1.430
	R^2	0.021	0.031	0.019
	Adj-R^2	0.011	0.022	0.009
	nobs	504	504	504
Hi	F	0.989	2.162	3.667
	R^2	0.015	0.023	0.034
	Adj-R^2	0.005	0.013	0.025
	nobs	504	504	504
Hi - Lo	F	15.879	1.924	18.752
	R^2	0.154	0.021	0.195
	Adj-R^2	0.146	0.011	0.187
	nobs	504	504	504
Avg	F	7.189	2.785	9.444
	R^2	0.056	0.033	0.062
	Adj-R^2	0.046	0.024	0.052
	nobs	504	504	504

Table 18: Fama-French *five* factor OLS regression on *value*-weighted portfolios sorted by FF5 R^2. *Contrast these results with Table 8.* We run the Fama-French 5 factor time series regression (14) onto value-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's FF5 R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
		coef	sd	t	coef	sd	t	coef	sd	t
FF5 R^2										
Lo	const	0.446	0.197	(2.261)	0.007	0.023	(0.298)	0.448	0.187	(2.397)
	Mkt-RF	23.440	4.817	(4.866)	1.283	0.567	(2.263)	22.165	4.486	(4.941)
	HML	23.340	7.358	(3.172)	1.328	1.062	(1.251)	21.973	7.170	(3.064)
	SMB	13.947	6.873	(2.029)	0.800	0.807	(0.991)	13.151	6.631	(1.983)
	CMA	-25.000	14.840	(-1.685)	-0.012	1.740	(-0.007)	-24.870	14.008	(-1.775)
	RMW	-4.564	9.839	(-0.464)	-1.230	0.736	(-1.671)	-3.475	9.603	(-0.362)
2	const	0.627	0.221	(2.842)	0.055	0.050	(1.106)	0.573	0.200	(2.858)
	Mkt-RF	17.542	5.733	(3.060)	1.736	1.308	(1.327)	16.097	5.197	(3.097)
	HML	13.231	9.775	(1.354)	-1.069	2.297	(-0.465)	14.663	8.758	(1.674)
	SMB	9.516	7.832	(1.215)	-2.162	1.609	(-1.344)	11.644	7.265	(1.603)
	CMA	-19.286	18.977	(-1.016)	2.233	3.481	(0.642)	-21.877	17.163	(-1.275)
	RMW	-14.465	10.481	(-1.380)	-3.153	2.688	(-1.173)	-11.326	9.324	(-1.215)
3	const	0.680	0.224	(3.033)	0.038	0.052	(0.734)	0.645	0.197	(3.276)
	Mkt-RF	13.574	5.188	(2.617)	3.395	1.754	(1.935)	10.309	4.467	(2.308)
	HML	10.930	9.135	(1.196)	2.295	3.244	(0.708)	8.764	8.421	(1.041)
	SMB	4.239	8.007	(0.529)	-0.278	1.665	(-0.167)	4.598	7.365	(0.624)
	CMA	-26.487	17.273	(-1.533)	-0.916	4.150	(-0.221)	-25.635	15.375	(-1.667)
	RMW	-2.957	11.536	(-0.256)	-2.886	2.421	(-1.192)	0.018	10.432	(0.002)
4	const	0.644	0.210	(3.071)	0.083	0.073	(1.133)	0.568	0.169	(3.359)
	Mkt-RF	11.541	5.644	(2.045)	5.734	2.530	(2.267)	5.733	4.565	(1.256)
	HML	7.313	10.594	(0.690)	5.125	5.798	(0.884)	2.060	10.577	(0.195)
	SMB	10.658	8.125	(1.312)	0.580	2.324	(0.250)	10.027	7.331	(1.368)
	CMA	-16.941	19.453	(-0.871)	-2.247	5.977	(-0.376)	-14.662	17.931	(-0.818)
	RMW	-0.282	10.846	(-0.026)	-1.856	2.852	(-0.651)	1.470	9.655	(0.152)
5	const	0.718	0.203	(3.544)	0.084	0.090	(0.939)	0.636	0.161	(3.949)
	Mkt-RF	11.136	5.098	(2.184)	7.464	3.110	(2.400)	3.712	3.939	(0.942)
	HML	6.141	9.487	(0.647)	4.293	5.969	(0.719)	1.873	9.194	(0.204)
	SMB	5.925	7.629	(0.777)	-0.482	2.369	(-0.204)	6.389	6.600	(0.968)
	CMA	-23.818	18.068	(-1.318)	0.182	6.752	(0.027)	-23.955	15.927	(-1.504)
	RMW	-12.667	10.057	(-1.260)	-3.978	3.637	(-1.094)	-8.671	8.217	(-1.055)
6	const	0.739	0.209	(3.537)	0.096	0.100	(0.958)	0.645	0.159	(4.068)
	Mkt-RF	9.365	5.736	(1.633)	7.625	3.300	(2.310)	1.775	3.725	(0.477)
	HML	15.101	9.541	(1.583)	4.956	6.429	(0.771)	10.158	8.011	(1.268)
	SMB	-2.987	7.490	(-0.399)	-1.584	2.732	(-0.580)	-1.419	6.103	(-0.232)
	CMA	-32.616	17.812	(-1.831)	-2.615	7.802	(-0.335)	-30.227	14.550	(-2.077)
	RMW	-10.244	10.612	(-0.965)	-2.213	3.781	(-0.585)	-8.048	8.818	(-0.913)
7	const	0.688	0.202	(3.408)	0.125	0.105	(1.185)	0.564	0.157	(3.581)
	Mkt-RF	6.713	5.278	(1.272)	7.950	3.331	(2.387)	-1.216	3.300	(-0.368)
	HML	8.344	8.814	(0.947)	2.449	5.169	(0.474)	5.932	7.721	(0.768)
	SMB	0.929	7.197	(0.129)	-1.593	2.983	(-0.534)	2.565	5.568	(0.461)
	CMA	-30.901	17.480	(-1.768)	1.185	7.116	(0.166)	-32.142	14.594	(-2.202)
	RMW	-7.822	9.193	(-0.851)	-3.933	4.179	(-0.941)	-3.877	7.241	(-0.535)
8	const	0.785	0.205	(3.836)	0.142	0.116	(1.221)	0.647	0.152	(4.256)
	Mkt-RF	3.440	5.695	(0.604)	8.874	3.613	(2.456)	-5.416	3.673	(-1.475)
	HML	9.448	8.914	(1.060)	3.690	6.624	(0.557)	5.695	7.889	(0.722)
	SMB	-1.554	7.584	(-0.205)	-1.815	3.435	(-0.528)	0.156	5.607	(0.028)
	CMA	-31.805	17.167	(-1.853)	-1.822	8.513	(-0.214)	-30.025	13.886	(-2.162)
	RMW	-14.284	8.757	(-1.631)	-8.192	4.491	(-1.824)	-6.152	6.917	(-0.889)
9	const	0.754	0.217	(3.484)	0.172	0.137	(1.249)	0.586	0.148	(3.966)
	Mkt-RF	4.006	6.159	(0.650)	9.652	4.221	(2.287)	-5.627	3.384	(-1.663)
	HML	8.566	9.172	(0.934)	1.761	7.209	(0.244)	6.840	6.681	(1.024)
	SMB	1.426	7.075	(0.202)	-1.649	4.086	(-0.404)	3.134	4.700	(0.667)
	CMA	-32.038	18.026	(-1.777)	-2.454	10.393	(-0.236)	-29.302	12.541	(-2.337)
	RMW	-8.404	9.197	(-0.914)	-8.676	5.207	(-1.666)	0.292	6.478	(0.045)
Hi	const	0.728	0.238	(3.056)	0.225	0.176	(1.280)	0.503	0.133	(3.785)
	Mkt-RF	-3.044	6.706	(-0.454)	9.020	5.079	(1.776)	-12.039	3.241	(-3.715)
	HML	-0.658	10.806	(-0.061)	0.896	9.567	(0.094)	-1.586	4.863	(-0.326)
	SMB	-0.148	8.244	(-0.018)	-4.360	6.117	(-0.713)	4.158	4.442	(0.936)
	CMA	-23.631	20.064	(-1.178)	-7.488	16.115	(-0.465)	-16.064	8.837	(-1.818)
	RMW	-18.151	10.172	(-1.784)	-16.017	7.864	(-2.037)	-2.143	5.184	(-0.413)
Hi - Lo	const	0.282	0.195	(1.446)	0.218	0.164	(1.329)	0.055	0.203	(0.271)
	Mkt-RF	-26.483	5.515	(-4.802)	7.736	4.825	(1.604)	-34.204	5.181	(-6.601)
	HML	-23.999	8.913	(-2.693)	-0.431	9.027	(-0.048)	-23.558	7.402	(-3.183)
	SMB	-14.095	7.910	(-1.782)	-5.160	5.898	(-0.875)	-8.993	7.310	(-1.230)
	CMA	1.369	13.713	(0.100)	-7.475	15.235	(-0.491)	8.806	13.088	(0.673)
	RMW	-13.587	11.127	(-1.221)	-14.787	7.474	(-1.978)	1.332	11.028	(0.121)
Avg	const	0.681	0.193	(3.520)	0.103	0.086	(1.195)	0.581	0.139	(4.181)
	Mkt-RF	9.771	5.116	(1.910)	6.273	2.718	(2.308)	3.549	3.291	(1.078)
	HML	10.176	8.515	(1.195)	2.572	5.010	(0.513)	7.637	6.785	(1.126)
	SMB	4.195	6.898	(0.608)	-1.254	2.503	(-0.501)	5.440	5.353	(1.016)
	CMA	-26.252	16.702	(-1.572)	-1.395	6.497	(-0.215)	-24.876	12.979	(-1.917)
	RMW	-9.384	8.983	(-1.045)	-5.213	3.325	(-1.568)	-4.191	6.935	(-0.604)

		Actual	**Ghost**	**Act - Gho**
FF5 R^2				
Lo	F	13.915	3.456	13.937
	R^2	0.108	0.029	0.107
	Adj-R^2	0.099	0.019	0.098
	nobs	504	504	504
2	F	6.334	1.093	8.013
	R^2	0.057	0.011	0.065
	Adj-R^2	0.048	0.001	0.056
	nobs	504	504	504
3	F	3.582	1.183	3.299
	R^2	0.036	0.022	0.033
	Adj-R^2	0.026	0.012	0.023
	nobs	504	504	504
4	F	2.530	1.685	1.622
	R^2	0.029	0.035	0.021
	Adj-R^2	0.019	0.025	0.011
	nobs	504	504	504
5	F	4.050	1.500	3.185
	R^2	0.040	0.038	0.033
	Adj-R^2	0.031	0.029	0.023
	nobs	504	504	504
6	F	2.311	1.213	1.934
	R^2	0.030	0.030	0.023
	Adj-R^2	0.020	0.021	0.013
	nobs	504	504	504
7	F	2.463	1.349	2.483
	R^2	0.026	0.030	0.025
	Adj-R^2	0.016	0.021	0.015
	nobs	504	504	504
8	F	2.561	2.351	2.482
	R^2	0.022	0.038	0.020
	Adj-R^2	0.012	0.028	0.010
	nobs	504	504	504
9	F	1.482	2.140	1.682
	R^2	0.019	0.034	0.019
	Adj-R^2	0.009	0.024	0.009
	nobs	504	504	504
Hi	F	1.601	2.077	3.364
	R^2	0.015	0.028	0.034
	Adj-R^2	0.005	0.018	0.024
	nobs	504	504	504
Hi - Lo	F	6.809	1.863	16.262
	R^2	0.096	0.026	0.143
	Adj-R^2	0.087	0.016	0.134
	nobs	504	504	504
Avg	F	3.371	1.866	2.688
	R^2	0.035	0.033	0.030
	Adj-R^2	0.025	0.024	0.020
	nobs	504	504	504

Table 19: Regressing the return differences between the elastic-net, FF3, and FF5 "Avg" value-weighted bins. For methods $m_1, m_2 \in \{\text{elastic-net}, \text{FF3}, \text{FF5}\}$ with $m_1 \neq m_2$, we compute the $(\mathsf{R}^2{}_{i,t-1})_{m_\ell}$ of each stock i at the end of month $t-1$, following the procedures described in Sections 2 and 5. For a given method m_ℓ, we sort the $(\mathsf{R}^2)_{m_\ell}$'s into decile bins, which necessarily means that at any month $t-1$, the set of stocks in the k-th bin B_{t-1}^{k,m_1} of method m_1, and the k-th bin B_{t-1}^{k,m_2} of method m_2 are necessarily different. We construct the "Actual", "Ghost" and "Act - Gho" value-weighted portfolio returns of bin k of method m_ℓ according to Section 2.2, and consider the returns of the "Avg" bin of method m_ℓ, which is the simple mean returns across bins $k = 10$ ("Hi") through $k = 1$ ("Lo"). Finally, we consider another difference in returns between the "Avg" bin of method m_1 and the "Avg" bin of method m_2. The take these difference-in-difference raw returns (so not excess returns) and regress them onto the Fama and French (2015) five-factors, plus the Jegadeesh and Titman (1993) and Carhart (1997) momentum factor, and the Pástor and Stambaugh (2003) liquidity factor. The parentheses show Newey and West (1987) robust t-statistics with 6 lags. The data sample is from December 1975 to December 2017.

	Actual '*Avg*'			**Ghost '*Avg*'**			**Act - Gho '*Avg*'**		
	(1) EN - FF3	(2) EN - FF5	(3) FF3 - FF5	(4) EN - FF3	(5) EN - FF5	(6) FF3 - FF5	(7) EN - FF3	(8) EN - FF5	(9) FF3 - FF5
Constant	0.001	0.0004	−0.0001	−0.0002	−0.0002	−0.00001	0.001	0.001	−0.0001
	(0.850)	(0.721)	(−0.460)	(−1.042)	(−1.229)	(−0.133)	(1.199)	(1.068)	(−0.460)
MktRF	−0.071	−0.073	−0.002	−0.015	−0.011	0.004	−0.056	−0.061	−0.006
	(−4.244)	(−4.154)	(−0.340)	(−3.206)	(−2.795)	(2.038)	(−3.319)	(−3.466)	(−1.466)
SMB	−0.022	−0.030	−0.008	−0.003	−0.006	−0.003	−0.019	−0.024	−0.005
	(−0.930)	(−1.135)	(−1.171)	(−0.556)	(−1.184)	(−1.413)	(−0.888)	(−0.998)	(−0.783)
HML	−0.014	−0.011	0.003	0.009	0.010	0.001	−0.023	−0.021	0.002
	(−0.458)	(−0.347)	(0.442)	(1.366)	(1.556)	(0.262)	(−0.848)	(−0.720)	(0.321)
CMA	−0.036	−0.040	−0.004	−0.030	−0.027	0.003	−0.006	−0.012	−0.006
	(−0.795)	(−0.845)	(−0.370)	(−2.538)	(−2.553)	(0.704)	(−0.158)	(−0.292)	(−0.625)
RMW	0.010	−0.011	−0.020	0.0001	−0.004	−0.004	0.010	−0.006	−0.016
	(0.234)	(−0.235)	(−2.259)	(0.008)	(−0.372)	(−1.758)	(0.287)	(−0.166)	(−2.098)
MOM	0.021	0.024	0.003	0.011	0.010	−0.001	0.009	0.013	0.004
	(1.577)	(1.817)	(0.914)	(3.054)	(2.609)	(−0.787)	(0.778)	(1.092)	(1.407)
LIQ	0.017	0.015	−0.002	−0.0004	0.001	0.001	0.018	0.014	−0.004
	(1.032)	(0.823)	(−0.525)	(−0.078)	(0.193)	(0.584)	(1.080)	(0.805)	(−0.801)
Observations	504	504	504	504	504	504	504	504	504
R^2	0.090	0.091	0.014	0.071	0.062	0.034	0.068	0.074	0.016
Adjusted R^2	0.078	0.078	0.001	0.058	0.049	0.021	0.055	0.061	0.002

Table 20: Regressing the return differences between the elastic-net, FF3, and FF5 "Hi - Lo" value-weighted bins. For methods $m_1, m_2 \in \{\text{elastic-net}, \text{FF3}, \text{FF5}\}$ with $m_1 \neq m_2$, we compute the $(\mathsf{R}^2{}_{i,t-1})_{m_\ell}$ of each stock i at the end of month $t-1$, following the procedures described in Sections 2 and 5. For a given method m_ℓ, we sort the $(\mathsf{R}^2)_{m_\ell}$'s into decile bins, which necessarily means that at any month $t-1$, the set of stocks in the k-th bin B_{t-1}^{k,m_1} of method m_1, and the k-th bin B_{t-1}^{k,m_2} of method m_2 are necessarily different. We construct the "Actual", "Ghost" and "Act - Gho" value-weighted portfolio returns of bin k of method m_ℓ according to Section 2.2, and consider the returns of the "Hi - Lo" bin of method m_ℓ, which is the difference in returns between the $k = 10$ ("Hi") bin and the $k = 1$ ("Lo") bin. Finally, we consider another difference in returns between the "Hi - Lo" bin of method m_1 and the "Hi - Lo" bin of method m_2. We take these difference-in-difference raw returns (so not excess returns) and regress them onto the Fama and French (2015) five-factors, plus the Jegadeesh and Titman (1993) and Carhart (1997) momentum factor, and the Pástor and Stambaugh (2003) liquidity factor. The parentheses show Newey and West (1987) robust t-statistics with 6 lags. The data sample is from December 1975 to December 2017.

	Actual '*Hi - Lo*'			**Ghost '*Hi - Lo*'**			**Act - Gho '*Hi - Lo*'**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	EN - FF3	EN - FF5	FF3 - FF5	EN - FF3	EN - FF5	FF3 - FF5	EN - FF3	EN - FF5	FF3 - FF5
Constant	−0.005	−0.006	−0.0002	0.002	0.001	−0.0001	−0.007	−0.007	−0.0001
	(−3.047)	(−3.257)	(−0.335)	(2.884)	(2.980)	(−0.328)	(−4.037)	(−4.142)	(−0.227)
MktRF	0.253	0.276	0.023	0.016	0.016	0.0003	0.237	0.261	0.023
	(5.181)	(5.799)	(1.377)	(1.215)	(1.446)	(0.044)	(5.125)	(5.720)	(1.716)
SMB	0.037	0.037	−0.0002	0.037	0.037	−0.001	−0.001	0.0003	0.001
	(0.520)	(0.570)	(−0.007)	(1.777)	(2.032)	(−0.061)	(−0.008)	(0.006)	(0.040)
HML	0.178	0.162	−0.016	0.020	0.006	−0.014	0.158	0.156	−0.002
	(2.268)	(2.235)	(−0.392)	(0.618)	(0.224)	(−0.644)	(2.114)	(2.081)	(−0.065)
CMA	0.160	0.141	−0.019	−0.047	−0.033	0.014	0.207	0.176	−0.031
	(1.330)	(1.231)	(−0.335)	(−1.109)	(−0.850)	(0.634)	(1.671)	(1.466)	(−0.594)
RMW	0.040	0.036	−0.004	0.0003	0.003	0.002	0.040	0.032	−0.008
	(0.468)	(0.455)	(−0.124)	(0.009)	(0.101)	(0.232)	(0.558)	(0.460)	(−0.296)
MOM	−0.014	−0.012	0.002	−0.023	−0.019	0.004	0.009	0.007	−0.002
	(−0.363)	(−0.354)	(0.100)	(−2.029)	(−2.045)	(0.521)	(0.267)	(0.221)	(−0.163)
LIQ	−0.079	−0.082	−0.003	−0.015	−0.015	0.0003	−0.065	−0.068	−0.003
	(−1.540)	(−1.663)	(−0.127)	(−1.026)	(−1.163)	(0.039)	(−1.269)	(−1.392)	(−0.170)
Observations	504	504	504	504	504	504	504	504	504
R^2	0.095	0.118	0.010	0.030	0.033	0.004	0.087	0.106	0.012
Adjusted R^2	0.083	0.106	−0.004	0.016	0.020	−0.010	0.074	0.093	−0.002

Table 21: Summary statistics of the elements sum of the FF3 and FF5 estimated coefficient vector by OLS R^2 bins. *Contrast these results against Table 1.* Note that unlike Table 1 where we show summary statistics for the number of non-zero elements in the estimated elastic-net coefficient, such analogous statistics for FFz would not be insightful because by nature of OLS, the number of non-zero elements in each estimated coefficient vector will generically be just z, for $z = 3, 5$. Actual stocks i are sorted by their elastic-net $\mathsf{R}^2{}_{i,t-1}$ at month $t-1$ into bins $\mathrm{Lo}, 2, ..., \mathrm{Hi}$ in ascending order. Each bin B^k_{t-1} then consists of a panel of estimated elastic-net coefficient vectors $\hat{\beta}_{i,t-1}$. Table (a) reports the summary statistics on the number of non-zero elements in the estimated coefficient vectors, not conditioning on time nor the cross-section. For instance in Table (a), the mean of the kth bin is calculated as $\frac{1}{T}\sum_t \frac{1}{|B^k_t|}\sum_{i\in B^k_t}(\#\text{ of non-zero elements in }\hat{\beta}_{i,t})$; other statistics are calculated analogously. Table (b) reports the summary statistics on the sum of elements in the estimated coefficient vectors, not conditioning on time nor the cross-section. For instance in Table (a), the mean of the kth bin is calculated as $\frac{1}{T}\sum_t \frac{1}{|B^k_t|}\sum_{i\in B^k_t}\mathbf{1}^\top\hat{\beta}_{i,t}$; other statistics are calculated analogously. Recall from Section 2.2 that for each actual stock i at month $t-1$, the ghost portfolio of stock i has $1-\mathbf{1}^\top\hat{\beta}_{i,t}$ proportion of wealth invested into the risk free asset. The columns "p5", "p25", "med", "p75" and "p95" respectively represent the 5th, 25th, 50th, 75th and 95th percentiles of the data. The sample period is from December 1975 to December 2017.

FF3 R^2	mean	std	min	max	p5	p25	med	p75	p95
Lo	0.552	1.365	-27.124	23.537	-1.267	-0.088	0.420	1.094	2.782
2	1.120	1.772	-36.890	34.919	-1.323	0.286	0.979	1.877	3.981
3	1.543	1.884	-44.690	48.485	-0.953	0.626	1.381	2.351	4.571
4	1.877	1.881	-31.623	52.756	-0.445	0.921	1.685	2.671	4.914
5	2.077	1.819	-31.525	49.430	-0.060	1.098	1.872	2.833	5.021
6	2.206	1.717	-28.203	47.369	0.157	1.220	1.991	2.949	5.032
7	2.278	1.597	-17.448	34.993	0.268	1.309	2.083	3.016	4.960
8	2.325	1.502	-13.892	43.418	0.309	1.379	2.157	3.078	4.952
9	2.280	1.480	-8.980	37.887	0.218	1.345	2.140	3.063	4.876
Hi	1.894	1.575	-9.797	11.837	-0.625	0.997	1.883	2.800	4.401

(a) Fama-French 3

FF5 R^2	mean	std	min	max	p5	p25	med	p75	p95
Lo	0.531	2.722	-34.460	48.679	-3.336	-0.643	0.402	1.616	4.744
2	0.912	3.339	-59.982	62.835	-3.909	-0.574	0.809	2.343	6.006
3	1.260	3.415	-61.032	57.593	-3.680	-0.273	1.181	2.750	6.409
4	1.624	3.320	-66.341	72.018	-3.059	0.165	1.528	3.029	6.514
5	1.882	3.091	-59.558	83.209	-2.399	0.497	1.759	3.205	6.408
6	2.030	2.880	-44.017	86.901	-2.012	0.695	1.925	3.306	6.305
7	2.115	2.611	-46.645	57.034	-1.696	0.830	2.048	3.364	6.068
8	2.173	2.402	-22.130	65.972	-1.534	0.929	2.128	3.417	5.971
9	2.048	2.301	-16.594	78.785	-1.685	0.849	2.052	3.318	5.661
Hi	1.456	2.452	-17.009	15.767	-2.837	0.326	1.654	2.853	5.033

(b) Fama-French 5

Table 22: Univariate equal- and value- weighted portfolio sort by elastic-net R^2, where each ghost portfolio must have at least one risky stock. *Contrast these results against Table 3.* We form equal- and value-weighted portfolios decile portfolios every month by regressing each stock's daily return over the past year onto every other stock using the elastic-net estimator (24). Stocks are sorted into deciles based on their elastic-net R^2 from the lowest (quantile 1, labelled "Lo") to highest (quantile 10, labelled "Hi"). The row labelled "Hi - Lo" is the monthly return difference between the "Hi" bin and the "Lo" bin. The row labelled "Avg" is the simple average of the monthly returns across the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' bins. The column labelled "Actual" reports the one month ahead portfolio returns of actual stocks (7). The column labelled "Ghost" reports the one month ahead returns of a portfolio that are constructed out of the estimated elastic-net beta coefficients according to (8). The column labelled "Act - Gho" reports the one month ahead returns of the portfolio of a long position in the actual stocks, and a short position in the corresponding ghost portfolios of the actual stocks. The value weight of the ghost portfolio of stock i is the same value weight as the actual stock i. For "Actual" and "Ghost", the columns labelled "mean" and "sd" are excess returns measured in monthly percentage terms (e.g. 1.0 means 1%). For "Act - Gho", the columns labelled "mean" and "sd" are total simple returns, and not excess returns, measured in monthly percentage terms. Robust Newey and West (1987) t-statistics are reported in column "t" in parentheses. The sample period is from December 1975 to December 2017.

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
EN R^2									
Lo	1.349	4.934	(4.664)	0.001	0.538	(0.421)	1.348	4.935	(4.664)
2	1.017	4.939	(3.354)	0.018	1.204	(1.195)	1.000	4.893	(3.363)
3	0.945	4.984	(3.126)	0.063	2.282	(1.111)	0.884	4.757	(3.205)
4	0.847	5.010	(2.919)	0.108	2.916	(1.079)	0.741	4.600	(3.083)
5	0.973	5.098	(3.384)	0.155	3.315	(1.178)	0.821	4.495	(3.741)
6	0.926	5.183	(3.250)	0.184	3.661	(1.202)	0.740	4.372	(3.703)
7	0.856	5.270	(2.999)	0.236	4.030	(1.309)	0.626	4.205	(3.429)
8	0.858	5.417	(3.011)	0.255	4.435	(1.235)	0.606	4.071	(3.749)
9	0.803	5.492	(2.747)	0.306	4.786	(1.301)	0.500	3.850	(3.359)
Hi	0.802	5.360	(2.916)	0.433	4.930	(1.707)	0.370	3.330	(3.401)
Hi - Lo	-0.547	4.082	(-3.486)	0.432	4.929	(1.708)	-0.978	4.758	(-3.422)
Avg	0.938	5.071	(3.350)	0.176	3.385	(1.394)	0.764	4.167	(4.031)

(a) Equal-weighted

	Actual			**Ghost**			**Act - Gho**		
	mean	sd	t	mean	sd	t	mean	sd	t
EN R^2									
Lo	0.899	4.344	(4.834)	-0.001	0.532	(-0.314)	0.906	4.346	(4.875)
2	0.646	4.288	(3.246)	0.017	1.227	(1.113)	0.634	4.252	(3.276)
3	0.534	4.266	(2.818)	0.025	1.776	(0.793)	0.508	4.177	(2.846)
4	0.696	4.164	(4.186)	0.049	2.150	(0.948)	0.649	4.008	(4.343)
5	0.686	4.219	(3.680)	0.082	2.580	(1.068)	0.605	3.944	(3.829)
6	0.739	4.365	(4.010)	0.135	2.920	(1.349)	0.605	3.957	(4.205)
7	0.675	4.465	(3.324)	0.125	3.276	(0.992)	0.551	3.935	(3.831)
8	0.567	4.550	(2.634)	0.165	3.622	(1.124)	0.404	3.760	(2.829)
9	0.633	4.740	(2.745)	0.195	4.005	(1.061)	0.440	3.679	(3.433)
Hi	0.669	4.818	(2.843)	0.336	4.364	(1.642)	0.335	3.341	(2.999)
Hi - Lo	-0.231	4.015	(-1.521)	0.337	4.363	(1.649)	-0.571	4.221	(-3.061)
Avg	0.674	4.236	(3.672)	0.113	2.782	(1.274)	0.564	3.679	(4.345)

(b) Value-weighted

Table 23: Fama-French *five* factor OLS regression on *equal*-weighted portfolios sorted by elastic-net R^2, where each ghost portfolio must have at least one risky stock. *Contrast these results against Table 7.* We run the Fama-French 5 factor time series regression (14) onto equal-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's elastic-net R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
		coef	sd	t	coef	sd	t	coef	sd	t
EN R^2										
Lo	const	1.178	0.248	(4.747)	-0.001	0.002	(-0.409)	1.179	0.248	(4.758)
	Mkt-RF	28.706	6.618	(4.337)	0.074	0.092	(0.812)	28.630	6.591	(4.344)
	HML	17.777	12.675	(1.403)	-0.297	0.178	(-1.667)	18.072	12.668	(1.427)
	SMB	17.359	8.873	(1.956)	0.137	0.092	(1.489)	17.221	8.864	(1.943)
	CMA	-30.479	23.513	(-1.296)	0.506	0.202	(2.505)	-30.989	23.497	(-1.319)
	RMW	-10.227	12.448	(-0.822)	0.205	0.192	(1.068)	-10.430	12.451	(-0.838)
2	const	0.815	0.255	(3.197)	0.012	0.014	(0.813)	0.804	0.249	(3.233)
	Mkt-RF	30.941	6.339	(4.881)	0.517	0.364	(1.420)	30.443	6.150	(4.950)
	HML	21.077	11.520	(1.830)	0.959	0.939	(1.021)	20.128	11.327	(1.777)
	SMB	21.035	8.679	(2.424)	0.479	0.395	(1.211)	20.570	8.599	(2.392)
	CMA	-32.148	20.469	(-1.571)	-0.509	1.208	(-0.421)	-31.675	20.118	(-1.574)
	RMW	-9.974	11.484	(-0.869)	-0.018	0.594	(-0.030)	-9.957	11.381	(-0.875)
3	const	0.732	0.262	(2.791)	0.038	0.054	(0.702)	0.697	0.235	(2.970)
	Mkt-RF	28.144	6.764	(4.161)	2.954	1.676	(1.763)	25.169	5.861	(4.294)
	HML	17.211	10.894	(1.580)	4.966	3.696	(1.344)	12.003	10.499	(1.143)
	SMB	21.629	8.343	(2.593)	0.716	1.446	(0.495)	20.885	7.987	(2.615)
	CMA	-26.738	20.938	(-1.277)	-2.820	4.250	(-0.664)	-23.694	19.622	(-1.208)
	RMW	-2.249	10.297	(-0.218)	-1.103	1.978	(-0.558)	-1.002	9.764	(-0.103)
4	const	0.728	0.260	(2.806)	0.066	0.097	(0.675)	0.665	0.207	(3.207)
	Mkt-RF	24.103	6.564	(3.672)	6.422	3.119	(2.059)	17.701	4.968	(3.563)
	HML	12.409	11.891	(1.044)	7.180	5.934	(1.210)	5.317	10.574	(0.503)
	SMB	14.883	8.960	(1.661)	-0.453	2.376	(-0.191)	15.279	7.969	(1.917)
	CMA	-31.255	21.601	(-1.447)	-3.835	6.709	(-0.572)	-27.553	19.104	(-1.442)
	RMW	-9.348	10.862	(-0.861)	-3.166	3.515	(-0.901)	-6.215	9.439	(-0.658)
5	const	0.895	0.265	(3.376)	0.098	0.128	(0.762)	0.799	0.192	(4.159)
	Mkt-RF	21.909	7.240	(3.026)	9.686	4.161	(2.328)	12.276	4.990	(2.460)
	HML	11.764	12.173	(0.966)	6.647	7.319	(0.908)	5.197	10.253	(0.507)
	SMB	10.185	9.382	(1.086)	-2.135	3.200	(-0.667)	12.323	7.864	(1.567)
	CMA	-32.259	23.211	(-1.390)	-2.282	8.734	(-0.261)	-29.927	19.042	(-1.572)
	RMW	-12.249	11.062	(-1.107)	-4.849	4.564	(-1.062)	-7.435	9.107	(-0.816)
6	const	0.900	0.269	(3.349)	0.126	0.150	(0.838)	0.772	0.179	(4.315)
	Mkt-RF	21.159	7.446	(2.842)	12.006	4.816	(2.493)	9.179	4.904	(1.872)
	HML	6.738	12.883	(0.523)	5.392	8.669	(0.622)	1.533	10.755	(0.143)
	SMB	5.309	10.068	(0.527)	-4.580	4.046	(-1.132)	9.968	7.989	(1.248)
	CMA	-35.921	24.798	(-1.449)	-0.209	11.203	(-0.019)	-35.904	18.954	(-1.894)
	RMW	-15.302	12.024	(-1.273)	-7.751	5.558	(-1.395)	-7.486	9.802	(-0.764)
7	const	0.879	0.276	(3.188)	0.192	0.175	(1.093)	0.692	0.171	(4.040)
	Mkt-RF	18.930	7.640	(2.478)	14.263	5.496	(2.595)	4.700	4.631	(1.015)
	HML	9.603	13.088	(0.734)	3.570	9.464	(0.377)	6.002	9.769	(0.614)
	SMB	-1.020	10.262	(-0.099)	-7.429	5.033	(-1.476)	6.443	7.413	(0.869)
	CMA	-37.036	26.835	(-1.380)	0.155	14.075	(0.011)	-37.053	18.080	(-2.049)
	RMW	-22.693	12.196	(-1.861)	-12.931	6.704	(-1.929)	-9.879	9.458	(-1.044)
8	const	0.908	0.278	(3.269)	0.243	0.197	(1.228)	0.668	0.157	(4.250)
	Mkt-RF	15.465	7.991	(1.935)	15.064	6.128	(2.458)	0.473	4.303	(0.110)
	HML	5.693	14.080	(0.404)	3.094	11.235	(0.275)	2.531	9.334	(0.271)
	SMB	-0.227	10.370	(-0.022)	-9.780	6.209	(-1.575)	9.416	6.725	(1.400)
	CMA	-36.091	28.472	(-1.268)	-4.371	17.960	(-0.243)	-31.801	16.451	(-1.933)
	RMW	-21.653	13.353	(-1.622)	-18.480	8.673	(-2.131)	-3.364	9.509	(-0.354)
9	const	0.898	0.295	(3.040)	0.320	0.225	(1.422)	0.582	0.157	(3.714)
	Mkt-RF	11.026	8.293	(1.330)	15.349	6.943	(2.211)	-4.335	3.516	(-1.233)
	HML	3.203	15.693	(0.204)	-0.176	12.893	(-0.014)	3.409	8.335	(0.409)
	SMB	-3.338	11.151	(-0.299)	-10.425	7.720	(-1.350)	7.041	6.312	(1.115)
	CMA	-33.389	30.504	(-1.095)	-4.407	22.114	(-0.199)	-29.157	15.245	(-1.913)
	RMW	-24.369	14.398	(-1.693)	-23.665	9.930	(-2.383)	-0.796	8.420	(-0.095)
Hi	const	0.894	0.287	(3.118)	0.445	0.249	(1.787)	0.451	0.118	(3.834)
	Mkt-RF	6.525	7.976	(0.818)	13.716	7.212	(1.902)	-7.187	2.552	(-2.816)
	HML	8.771	15.439	(0.568)	-1.114	14.586	(-0.076)	9.791	5.595	(1.750)
	SMB	-1.711	11.050	(-0.155)	-1.993	9.073	(-0.220)	0.350	3.937	(0.089)
	CMA	-35.196	28.736	(-1.225)	-10.767	23.599	(-0.456)	-24.314	10.825	(-2.246)
	RMW	-19.238	14.218	(-1.353)	-19.723	11.591	(-1.702)	0.338	4.948	(0.068)
Hi - Lo	const	-0.284	0.162	(-1.752)	0.446	0.248	(1.796)	-0.728	0.238	(-3.056)
	Mkt-RF	-22.181	3.595	(-6.170)	13.641	7.178	(1.901)	-35.817	6.212	(-5.765)
	HML	-9.006	8.002	(-1.125)	-0.817	14.554	(-0.056)	-8.281	12.505	(-0.662)
	SMB	-19.070	6.913	(-2.759)	-2.130	9.051	(-0.235)	-16.871	7.912	(-2.132)
	CMA	-4.717	13.647	(-0.346)	-11.274	23.552	(-0.479)	6.674	19.932	(0.335)
	RMW	-9.011	9.027	(-0.998)	-19.928	11.540	(-1.727)	10.768	11.058	(0.974)
Avg	const	0.883	0.259	(3.408)	0.154	0.122	(1.257)	0.731	0.170	(4.304)
	Mkt-RF	20.691	7.035	(2.941)	9.005	3.835	(2.348)	11.705	4.236	(2.763)
	HML	11.425	12.367	(0.924)	3.022	7.076	(0.427)	8.398	8.605	(0.976)
	SMB	8.410	9.258	(0.908)	-3.546	3.643	(-0.973)	11.949	6.787	(1.761)
	CMA	-33.051	24.055	(-1.374)	-2.854	10.269	(-0.278)	-30.207	16.713	(-1.807)
	RMW	-14.730	11.500	(-1.281)	-9.148	4.767	(-1.919)	-5.623	8.569	(-0.656)

		Actual	**Ghost**	**Act - Gho**
EN R^2				
Lo	F	11.134	2.144	11.200
	R^2	0.094	0.017	0.094
	Adj-R^2	0.085	0.007	0.085
	nobs	504	504	504
2	F	13.611	0.677	14.491
	R^2	0.110	0.011	0.111
	Adj-R^2	0.101	0.001	0.102
	nobs	504	504	504
3	F	10.953	0.681	12.832
	R^2	0.084	0.020	0.085
	Adj-R^2	0.075	0.011	0.076
	nobs	504	504	504
4	F	9.917	1.073	10.899
	R^2	0.070	0.028	0.069
	Adj-R^2	0.061	0.018	0.059
	nobs	504	504	504
5	F	7.861	1.516	8.399
	R^2	0.057	0.033	0.053
	Adj-R^2	0.047	0.023	0.043
	nobs	504	504	504
6	F	7.568	2.071	7.126
	R^2	0.057	0.034	0.059
	Adj-R^2	0.047	0.024	0.050
	nobs	504	504	504
7	F	6.465	2.672	4.869
	R^2	0.048	0.036	0.045
	Adj-R^2	0.038	0.027	0.035
	nobs	504	504	504
8	F	4.474	3.148	2.678
	R^2	0.037	0.035	0.033
	Adj-R^2	0.027	0.025	0.023
	nobs	504	504	504
9	F	2.799	3.197	2.068
	R^2	0.027	0.033	0.023
	Adj-R^2	0.018	0.024	0.013
	nobs	504	504	504
Hi	F	1.552	2.716	2.964
	R^2	0.018	0.029	0.025
	Adj-R^2	0.008	0.019	0.015
	nobs	504	504	504
Hi - Lo	F	12.630	2.726	13.091
	R^2	0.088	0.029	0.131
	Adj-R^2	0.079	0.019	0.122
	nobs	504	504	504
Avg	F	7.138	2.785	9.366
	R^2	0.055	0.033	0.061
	Adj-R^2	0.046	0.024	0.052
	nobs	504	504	504

Table 24: Fama-French *five* factor OLS regression on *value*-weighted portfolios sorted by elastic-net R^2, where each ghost portfolio must have at least one risky stock. *Contrast these results against Table 8.* We run the Fama-French 5 factor time series regression (14) onto value-weighted portfolios, where stocks in the portfolios are sorted into deciles by each stock's elastic-net R^2. The "Actual" column refers to the portfolio of actual stocks as per (10), the "Ghost" column refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" column refers to the portfolio of long-short actuals against ghost stocks as per (12). The row "Hi - Lo" refers to the return difference between the "Hi" portfolio bin less the "Lo" portfolio bin. The row "Avg" refers to the simple average of returns across the ten $k =$ 'Lo', $2, \ldots,$ 'Hi' portfolio bins. The left table shows the regression coefficient estimates and test statistics, while the right table shows aggregate statistics of the regression. The columns "coef" are the OLS coefficient estimates and "sd" are the standard errors, both are multiplied by 100. The column "t" show robust Newey and West (1987) t-statistic of the estimated coefficients. The row labels on the right table are the F test statistic, R^2 and adjusted R^2 in decimals, and number of time observations in the regression. The sample period is from December 1975 to December 2017.

		Actual			**Ghost**			**Act - Gho**		
		coef	sd	t	coef	sd	t	coef	sd	t
EN R^2										
Lo	const	1.051	0.205	(5.120)	-0.003	0.002	(-1.104)	1.061	0.205	(5.180)
	Mkt-RF	-2.007	5.545	(-0.362)	0.045	0.091	(0.494)	-2.079	5.514	(-0.377)
	HML	6.369	8.813	(0.723)	-0.343	0.168	(-2.041)	6.724	8.824	(0.762)
	SMB	11.222	6.448	(1.740)	0.123	0.090	(1.368)	11.098	6.434	(1.725)
	CMA	-34.585	17.382	(-1.990)	0.557	0.196	(2.839)	-35.163	17.383	(-2.023)
	RMW	-7.206	9.512	(-0.758)	0.185	0.184	(1.002)	-7.401	9.536	(-0.776)
2	const	0.845	0.197	(4.289)	-0.000	0.004	(-0.109)	0.849	0.197	(4.320)
	Mkt-RF	2.589	5.833	(0.444)	0.192	0.126	(1.527)	2.389	5.777	(0.414)
	HML	5.342	10.278	(0.520)	-0.043	0.299	(-0.144)	5.410	10.157	(0.533)
	SMB	3.613	5.891	(0.613)	0.290	0.154	(1.889)	3.323	5.835	(0.570)
	CMA	-19.499	14.823	(-1.315)	0.373	0.326	(1.143)	-19.895	14.754	(-1.349)
	RMW	-5.624	7.398	(-0.760)	0.287	0.218	(1.314)	-5.926	7.408	(-0.800)
3	const	0.553	0.192	(2.880)	0.021	0.019	(1.103)	0.536	0.188	(2.848)
	Mkt-RF	5.543	5.002	(1.108)	0.619	0.587	(1.054)	4.930	4.777	(1.032)
	HML	16.087	7.395	(2.175)	1.930	1.511	(1.277)	14.292	7.307	(1.956)
	SMB	6.673	6.251	(1.067)	0.773	0.642	(1.203)	5.992	6.216	(0.964)
	CMA	-27.938	14.425	(-1.937)	-1.717	1.775	(-0.967)	-26.425	14.300	(-1.848)
	RMW	-0.044	7.010	(-0.006)	-0.217	0.702	(-0.308)	0.163	6.999	(0.023)
4	const	0.559	0.188	(2.971)	0.033	0.043	(0.777)	0.525	0.175	(2.996)
	Mkt-RF	4.820	4.931	(0.977)	2.597	1.653	(1.571)	2.339	4.423	(0.529)
	HML	7.968	8.112	(0.982)	5.161	3.681	(1.402)	2.846	9.408	(0.302)
	SMB	1.624	5.837	(0.278)	0.519	1.197	(0.433)	1.121	5.582	(0.201)
	CMA	-19.563	14.918	(-1.311)	-3.366	3.634	(-0.926)	-16.251	15.408	(-1.055)
	RMW	1.884	7.818	(0.241)	-0.971	1.511	(-0.643)	2.849	7.739	(0.368)
5	const	0.664	0.172	(3.858)	0.026	0.062	(0.426)	0.640	0.149	(4.300)
	Mkt-RF	4.174	4.907	(0.851)	4.349	2.308	(1.884)	-0.169	3.896	(-0.043)
	HML	10.277	7.977	(1.288)	5.214	4.284	(1.217)	5.077	8.504	(0.597)
	SMB	0.120	5.418	(0.022)	0.086	1.668	(0.051)	0.039	4.950	(0.008)
	CMA	-30.948	13.827	(-2.238)	-2.872	4.363	(-0.658)	-28.094	13.333	(-2.107)
	RMW	-5.892	6.436	(-0.915)	-1.447	2.180	(-0.664)	-4.435	6.034	(-0.735)
6	const	0.605	0.184	(3.289)	0.058	0.087	(0.673)	0.547	0.150	(3.651)
	Mkt-RF	5.342	4.853	(1.101)	6.067	2.923	(2.076)	-0.710	3.377	(-0.210)
	HML	9.819	7.704	(1.275)	4.915	5.555	(0.885)	4.918	7.849	(0.627)
	SMB	-3.810	5.971	(-0.638)	-1.571	2.233	(-0.704)	-2.261	4.888	(-0.462)
	CMA	-33.107	15.091	(-2.194)	-1.630	6.389	(-0.255)	-31.480	13.977	(-2.252)
	RMW	-9.131	7.569	(-1.206)	-3.207	3.152	(-1.017)	-5.916	6.646	(-0.890)
7	const	0.881	0.199	(4.428)	0.104	0.107	(0.975)	0.779	0.152	(5.130)
	Mkt-RF	3.335	5.246	(0.636)	8.358	3.536	(2.364)	-5.016	3.663	(-1.369)
	HML	7.403	8.163	(0.907)	5.171	6.581	(0.786)	2.279	7.443	(0.306)
	SMB	-0.876	6.897	(-0.127)	-2.101	2.984	(-0.704)	1.206	5.515	(0.219)
	CMA	-37.740	15.717	(-2.401)	-2.204	7.756	(-0.284)	-35.551	12.814	(-2.774)
	RMW	-21.979	8.240	(-2.667)	-6.004	3.906	(-1.537)	-16.017	6.743	(-2.375)
8	const	0.765	0.211	(3.618)	0.151	0.134	(1.133)	0.616	0.153	(4.029)
	Mkt-RF	0.429	6.197	(0.069)	8.949	3.809	(2.350)	-8.512	3.940	(-2.160)
	HML	6.725	9.302	(0.723)	3.024	6.815	(0.444)	3.651	6.592	(0.554)
	SMB	-2.678	7.605	(-0.352)	-6.050	4.015	(-1.507)	3.341	5.595	(0.597)
	CMA	-30.191	16.697	(-1.808)	-3.348	9.727	(-0.344)	-26.866	11.877	(-2.262)
	RMW	-18.524	8.893	(-2.083)	-10.106	4.970	(-2.033)	-8.479	6.743	(-1.257)
9	const	0.772	0.229	(3.369)	0.185	0.165	(1.123)	0.587	0.139	(4.212)
	Mkt-RF	-0.651	6.277	(-0.104)	9.545	4.793	(1.991)	-10.183	3.214	(-3.169)
	HML	8.108	10.021	(0.809)	3.569	9.019	(0.396)	4.517	5.744	(0.786)
	SMB	-2.058	7.531	(-0.273)	-7.390	5.239	(-1.411)	5.344	4.322	(1.236)
	CMA	-37.486	19.307	(-1.942)	-10.763	14.142	(-0.761)	-26.710	10.803	(-2.472)
	RMW	-14.475	9.786	(-1.479)	-16.601	6.880	(-2.413)	2.152	5.534	(0.389)
Hi	const	0.740	0.249	(2.967)	0.341	0.197	(1.733)	0.400	0.128	(3.121)
	Mkt-RF	-0.703	7.037	(-0.100)	9.594	5.449	(1.761)	-10.245	3.284	(-3.119)
	HML	-1.089	11.922	(-0.091)	0.892	10.818	(0.082)	-1.990	4.995	(-0.398)
	SMB	1.260	8.571	(0.147)	-1.570	6.880	(-0.228)	2.794	4.232	(0.660)
	CMA	-19.585	21.712	(-0.902)	-11.149	17.608	(-0.633)	-8.295	8.702	(-0.953)
	RMW	-16.738	10.694	(-1.565)	-14.841	8.939	(-1.660)	-1.879	5.013	(-0.375)
Hi - Lo	const	-0.311	0.160	(-1.944)	0.344	0.196	(1.751)	-0.661	0.187	(-3.539)
	Mkt-RF	1.304	4.848	(0.269)	9.549	5.416	(1.763)	-8.166	4.578	(-1.784)
	HML	-7.458	10.240	(-0.728)	1.235	10.800	(0.114)	-8.714	8.036	(-1.084)
	SMB	-9.962	5.427	(-1.836)	-1.693	6.858	(-0.247)	-8.303	5.636	(-1.473)
	CMA	14.999	14.358	(1.045)	-11.706	17.570	(-0.666)	26.868	14.266	(1.883)
	RMW	-9.531	8.262	(-1.154)	-15.026	8.903	(-1.688)	5.522	9.407	(0.587)
Avg	const	0.744	0.187	(3.985)	0.092	0.077	(1.199)	0.654	0.142	(4.606)
	Mkt-RF	2.287	5.182	(0.441)	5.032	2.406	(2.092)	-2.726	3.577	(-0.762)
	HML	7.701	7.589	(1.015)	2.949	4.472	(0.659)	4.772	5.920	(0.806)
	SMB	1.509	6.039	(0.250)	-1.689	2.268	(-0.745)	3.200	4.658	(0.687)
	CMA	-29.064	14.597	(-1.991)	-3.612	5.920	(-0.610)	-25.473	11.462	(-2.222)
	RMW	-9.773	7.111	(-1.374)	-5.292	2.840	(-1.864)	-4.489	5.643	(-0.795)

		Actual	**Ghost**	**Act - Gho**
EN R^2				
Lo	F	3.227	2.333	3.238
	R^2	0.025	0.020	0.026
	Adj-R^2	0.016	0.010	0.016
	nobs	504	504	504
2	F	1.509	2.574	1.527
	R^2	0.010	0.012	0.010
	Adj-R^2	0.000	0.002	0.000
	nobs	504	504	504
3	F	2.053	0.436	1.883
	R^2	0.019	0.015	0.017
	Adj-R^2	0.010	0.005	0.007
	nobs	504	504	504
4	F	1.097	0.530	0.708
	R^2	0.010	0.032	0.007
	Adj-R^2	0.000	0.022	-0.003
	nobs	504	504	504
5	F	3.067	0.881	2.448
	R^2	0.022	0.033	0.019
	Adj-R^2	0.013	0.023	0.009
	nobs	504	504	504
6	F	2.060	1.203	1.873
	R^2	0.026	0.028	0.024
	Adj-R^2	0.017	0.019	0.014
	nobs	504	504	504
7	F	4.624	1.968	4.729
	R^2	0.042	0.036	0.042
	Adj-R^2	0.032	0.026	0.033
	nobs	504	504	504
8	F	1.872	2.202	2.407
	R^2	0.020	0.032	0.022
	Adj-R^2	0.010	0.023	0.012
	nobs	504	504	504
9	F	1.699	2.580	3.156
	R^2	0.019	0.036	0.028
	Adj-R^2	0.009	0.026	0.018
	nobs	504	504	504
Hi	F	1.258	2.104	2.137
	R^2	0.013	0.027	0.026
	Adj-R^2	0.003	0.017	0.016
	nobs	504	504	504
Hi - Lo	F	0.977	2.123	3.236
	R^2	0.010	0.027	0.037
	Adj-R^2	-0.000	0.017	0.027
	nobs	504	504	504
Avg	F	2.484	2.096	2.149
	R^2	0.021	0.032	0.018
	Adj-R^2	0.011	0.022	0.008
	nobs	504	504	504

Table 25: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *total volatility*. We control for the effects of total volatility. At the end of each month, we sort the actual stocks' total volatility into 5-tiles. Total volatility of actual stock i at month $t-1$ is the standard deviation of past one year's daily returns leading up until month $t-1$. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k = $ 'Lo'$, 2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo TotalVol	0.801	0.761	0.730	0.719	0.739	0.711	0.706	0.778	0.693	0.737	-0.064	-0.072
		(4.688)	(4.801)	(4.400)	(4.477)	(4.288)	(4.273)	(4.173)	(4.224)	(3.852)	(4.105)	(-0.496)	(-0.561)
	2	0.986	0.835	0.763	0.672	0.680	0.842	0.651	0.772	0.542	0.771	-0.215	-0.097
		(4.104)	(3.453)	(3.328)	(3.109)	(2.960)	(3.554)	(2.753)	(2.977)	(2.101)	(2.993)	(-1.095)	(-0.503)
	3	1.309	0.889	0.785	0.895	0.500	0.728	0.722	0.717	0.763	0.790	-0.519	-0.438
		(4.790)	(3.124)	(2.768)	(3.039)	(1.593)	(2.589)	(2.299)	(2.337)	(2.316)	(2.287)	(-1.758)	(-1.279)
	4	1.107	0.832	0.685	0.289	0.290	0.314	0.536	0.661	0.585	0.190	-0.917	-0.752
		(3.034)	(2.444)	(1.963)	(0.906)	(0.743)	(0.783)	(1.358)	(1.589)	(1.341)	(0.424)	(-3.194)	(-2.395)
	Hi TotalVol	0.690	0.278	0.178	-0.045	-0.138	-0.450	-0.161	-0.216	-0.170	-0.138	-0.828	-0.644
		(1.472)	(0.648)	(0.360)	(-0.094)	(-0.263)	(-0.890)	(-0.325)	(-0.410)	(-0.314)	(-0.231)	(-1.838)	(-1.449)
	Avg	0.979	0.719	0.628	0.506	0.414	0.429	0.491	0.543	0.482	0.470	-0.508	-0.401
		(3.743)	(2.912)	(2.472)	(2.040)	(1.488)	(1.562)	(1.746)	(1.828)	(1.533)	(1.413)	(-2.516)	(-1.859)
Ghost	Lo TotalVol	0.002	0.032	0.059	0.038	0.086	0.109	0.123	0.174	0.220	0.368	0.366	0.311
		(0.378)	(1.491)	(1.376)	(0.723)	(1.281)	(1.360)	(1.395)	(1.512)	(1.731)	(2.568)	(2.573)	(2.282)
	2	-0.001	0.042	0.070	0.132	0.136	0.194	0.234	0.242	0.239	0.432	0.433	0.367
		(-0.390)	(1.344)	(1.185)	(1.305)	(1.216)	(1.490)	(1.581)	(1.528)	(1.247)	(1.959)	(1.968)	(1.720)
	3	0.000	0.018	0.075	0.110	0.153	0.186	0.216	0.202	0.233	0.332	0.331	0.223
		(0.158)	(1.162)	(1.295)	(1.158)	(1.214)	(1.164)	(1.180)	(0.992)	(0.947)	(1.142)	(1.142)	(0.779)
	4	-0.001	0.013	0.030	0.059	0.116	0.134	0.142	0.145	0.407	0.319	0.320	0.239
		(-0.389)	(1.233)	(1.160)	(0.671)	(0.843)	(0.758)	(0.647)	(0.585)	(1.357)	(0.879)	(0.883)	(0.656)
	Hi TotalVol	0.000	0.000	0.004	0.027	0.073	0.047	0.082	0.075	0.152	0.211	0.211	0.079
		(0.047)	(0.145)	(0.759)	(1.139)	(1.129)	(0.442)	(0.421)	(0.297)	(0.454)	(0.457)	(0.457)	(0.173)
	Avg	0.000	0.021	0.048	0.073	0.113	0.134	0.159	0.168	0.250	0.332	0.332	0.244
		(0.028)	(1.517)	(1.350)	(1.083)	(1.178)	(1.079)	(1.007)	(0.891)	(1.077)	(1.172)	(1.174)	(0.879)
Act - Gho	Lo TotalVol	0.802	0.733	0.671	0.681	0.653	0.603	0.583	0.605	0.474	0.370	-0.432	-0.386
		(4.702)	(4.667)	(4.126)	(4.620)	(4.130)	(3.955)	(4.064)	(4.213)	(3.740)	(3.001)	(-2.842)	(-2.770)
	2	0.996	0.793	0.691	0.541	0.544	0.649	0.421	0.533	0.303	0.343	-0.654	-0.470
		(4.154)	(3.358)	(3.251)	(2.870)	(2.932)	(3.590)	(2.356)	(2.910)	(1.879)	(2.809)	(-2.838)	(-2.280)
	3	1.322	0.879	0.716	0.786	0.349	0.545	0.508	0.510	0.538	0.463	-0.859	-0.671
		(4.846)	(3.137)	(2.734)	(3.074)	(1.346)	(2.545)	(2.257)	(2.513)	(2.633)	(3.247)	(-2.877)	(-2.293)
	4	1.123	0.827	0.661	0.236	0.177	0.190	0.398	0.526	0.181	-0.124	-1.248	-1.003
		(3.087)	(2.445)	(1.942)	(0.783)	(0.531)	(0.609)	(1.402)	(1.837)	(0.731)	(-0.619)	(-3.659)	(-3.278)
	Hi TotalVol	0.701	0.291	0.180	-0.064	-0.208	-0.495	-0.242	-0.297	-0.319	-0.362	-1.063	-0.747
		(1.495)	(0.679)	(0.366)	(-0.136)	(-0.420)	(-1.070)	(-0.622)	(-0.735)	(-0.957)	(-1.415)	(-2.462)	(-1.881)
	Avg	0.989	0.705	0.584	0.436	0.303	0.298	0.334	0.375	0.235	0.138	-0.851	-0.655
		(3.789)	(2.905)	(2.449)	(2.032)	(1.349)	(1.458)	(1.778)	(2.043)	(1.454)	(1.126)	(-3.599)	(-3.267)

Table 26: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *idiosyncratic volatility.* We control for the effects of idiosyncratic volatility. At the end of each month, we sort all the actual stocks' idiosyncratic volatility into 5-tiles. Total volatility of actual stock i at month $t-1$ is the standard deviation of past one year's daily returns leading up until month $t-1$. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3* α
Actual	Lo IdioVol	0.833	0.681	0.732	0.804	0.667	0.689	0.725	0.696	0.675	0.690	-0.143	-0.168
		(5.192)	(4.179)	(4.428)	(4.767)	(3.784)	(4.012)	(3.844)	(3.530)	(3.731)	(3.327)	(-0.893)	(-0.994)
	2	1.026	0.853	0.756	0.751	0.744	0.683	0.822	0.844	0.499	0.910	-0.116	-0.035
		(4.427)	(3.648)	(3.571)	(3.397)	(3.195)	(2.930)	(3.261)	(3.165)	(1.736)	(3.179)	(-0.534)	(-0.157)
	3	1.165	0.813	0.812	0.752	0.602	0.722	0.607	0.652	0.871	0.677	-0.489	-0.396
		(4.416)	(2.884)	(2.885)	(2.769)	(1.985)	(2.469)	(1.873)	(1.991)	(2.456)	(1.704)	(-1.508)	(-1.099)
	4	1.328	0.751	0.579	0.400	0.046	0.253	0.436	0.768	0.540	0.051	-1.277	-1.173
		(3.786)	(2.235)	(1.690)	(1.179)	(0.133)	(0.657)	(1.077)	(1.825)	(1.286)	(0.108)	(-4.536)	(-4.097)
	Hi IdioVol	0.699	0.023	0.256	-0.300	0.199	-0.148	-0.291	0.253	-0.404	-0.000	-0.699	-0.599
		(1.509)	(0.057)	(0.505)	(-0.670)	(0.411)	(-0.272)	(-0.595)	(0.470)	(-0.806)	(-0.000)	(-1.595)	(-1.431)
	Avg	1.010	0.624	0.627	0.481	0.451	0.440	0.460	0.643	0.436	0.466	-0.545	-0.474
		(3.941)	(2.588)	(2.430)	(2.000)	(1.708)	(1.577)	(1.579)	(2.074)	(1.378)	(1.298)	(-2.460)	(-2.082)
Ghost	Lo IdioVol	0.032	0.045	0.060	0.099	0.101	0.158	0.190	0.202	0.285	0.366	0.333	0.281
		(2.002)	(1.054)	(1.029)	(1.272)	(1.177)	(1.497)	(1.510)	(1.489)	(1.917)	(2.097)	(1.977)	(1.723)
	2	0.013	0.060	0.081	0.095	0.121	0.249	0.210	0.235	0.222	0.453	0.440	0.335
		(0.992)	(1.232)	(0.986)	(0.998)	(1.162)	(1.715)	(1.297)	(1.358)	(1.058)	(1.803)	(1.780)	(1.340)
	3	0.000	0.025	0.093	0.104	0.139	0.161	0.234	0.167	0.304	0.425	0.424	0.333
		(0.127)	(1.445)	(1.636)	(1.165)	(1.126)	(1.032)	(1.271)	(0.803)	(1.168)	(1.340)	(1.341)	(1.070)
	4	-0.001	0.004	0.035	0.047	0.038	0.103	0.146	0.113	0.235	0.378	0.379	0.268
		(-0.469)	(1.299)	(1.781)	(1.179)	(0.424)	(0.713)	(0.736)	(0.474)	(0.797)	(0.917)	(0.920)	(0.696)
	Hi IdioVol	-0.000	0.001	-0.001	0.016	0.022	0.051	0.015	0.041	0.121	0.233	0.233	0.033
		(-0.221)	(0.419)	(-0.163)	(1.124)	(0.786)	(0.716)	(0.132)	(0.205)	(0.406)	(0.514)	(0.515)	(0.079)
	Avg	0.009	0.027	0.054	0.072	0.084	0.145	0.159	0.152	0.233	0.371	0.362	0.250
		(1.519)	(1.281)	(1.330)	(1.226)	(1.045)	(1.220)	(1.071)	(0.837)	(0.995)	(1.198)	(1.179)	(0.854)
Act - Gho	Lo IdioVol	0.806	0.638	0.672	0.706	0.566	0.531	0.536	0.494	0.392	0.326	-0.481	-0.452
		(5.072)	(4.079)	(4.388)	(4.548)	(3.802)	(3.580)	(3.576)	(3.603)	(3.230)	(2.609)	(-3.274)	(-3.262)
	2	1.023	0.793	0.676	0.657	0.623	0.434	0.617	0.613	0.279	0.461	-0.562	-0.376
		(4.416)	(3.527)	(3.623)	(3.325)	(3.481)	(2.400)	(3.483)	(3.240)	(1.611)	(3.736)	(-2.552)	(-1.913)
	3	1.180	0.797	0.723	0.650	0.463	0.565	0.374	0.484	0.560	0.260	-0.919	-0.739
		(4.480)	(2.863)	(2.768)	(2.720)	(1.850)	(2.422)	(1.660)	(2.345)	(2.717)	(1.546)	(-3.274)	(-2.706)
	4	1.345	0.757	0.549	0.356	0.012	0.157	0.298	0.665	0.307	-0.321	-1.666	-1.450
		(3.842)	(2.254)	(1.626)	(1.095)	(0.037)	(0.486)	(0.987)	(2.348)	(1.172)	(-1.618)	(-5.097)	(-4.974)
	Hi IdioVol	0.710	0.036	0.263	-0.304	0.182	-0.199	-0.304	0.210	-0.529	-0.241	-0.951	-0.654
		(1.533)	(0.088)	(0.520)	(-0.685)	(0.385)	(-0.391)	(-0.711)	(0.496)	(-1.593)	(-0.923)	(-2.303)	(-1.661)
	Avg	1.013	0.604	0.577	0.413	0.369	0.298	0.304	0.493	0.202	0.097	-0.916	-0.734
		(3.969)	(2.570)	(2.421)	(1.939)	(1.717)	(1.416)	(1.535)	(2.585)	(1.222)	(0.784)	(-4.138)	(-3.870)

Table 27: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *FF3* R^2. We control for the effects of FF3 R^2. At the end of each month, we sort all the actual stock's $\mathsf{R}^2{}_{\mathrm{FF3},i,t-1}$ value into 5-tiles, which is obtained from regressing the past daily returns onto the Fama-French 3-factors according to the procedure of Section 5. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \dots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo FF3 R^2	0.924	0.839	0.799	0.615	0.533	0.301	0.622	0.722	0.446	0.544	-0.381	-0.444
		(3.699)	(3.293)	(3.305)	(2.335)	(2.037)	(1.187)	(2.398)	(2.747)	(1.641)	(1.735)	(-1.715)	(-1.899)
	2	1.128	0.886	0.541	0.709	0.530	0.250	0.708	0.754	0.791	0.515	-0.613	-0.662
		(4.827)	(3.704)	(2.286)	(2.683)	(2.087)	(1.014)	(2.829)	(2.798)	(2.687)	(2.298)	(-3.230)	(-3.424)
	3	0.949	0.794	0.598	0.719	0.588	0.497	0.678	0.799	0.601	0.576	-0.373	-0.435
		(4.486)	(3.682)	(2.603)	(3.246)	(2.545)	(2.400)	(2.890)	(3.454)	(2.492)	(2.203)	(-1.959)	(-2.001)
	4	0.838	0.730	0.845	0.761	0.592	0.684	0.734	0.769	0.450	0.689	-0.149	-0.165
		(4.524)	(3.729)	(4.055)	(3.740)	(2.637)	(2.995)	(3.304)	(3.266)	(1.756)	(2.929)	(-0.801)	(-0.853)
	Hi FF3 R^2	0.761	0.793	0.664	0.481	0.737	0.544	0.603	0.567	0.546	0.684	-0.077	-0.102
		(4.213)	(4.010)	(3.117)	(1.968)	(3.085)	(2.301)	(2.337)	(2.244)	(2.152)	(2.792)	(-0.483)	(-0.621)
	Avg	0.920	0.808	0.689	0.657	0.596	0.455	0.669	0.722	0.567	0.602	-0.319	-0.361
		(4.971)	(4.180)	(3.522)	(3.165)	(2.802)	(2.207)	(3.087)	(3.281)	(2.454)	(2.748)	(-2.880)	(-3.258)
Ghost	Lo FF3 R^2	-0.001	-0.000	-0.000	-0.001	-0.001	0.003	0.012	0.015	0.021	0.185	0.186	0.088
		(-0.262)	(-0.204)	(-0.143)	(-0.372)	(-0.397)	(0.891)	(1.762)	(1.133)	(0.745)	(1.333)	(1.339)	(0.681)
	2	-0.001	0.002	0.010	0.048	0.050	0.061	0.059	0.098	0.093	0.266	0.266	0.191
		(-0.338)	(0.356)	(0.909)	(1.560)	(1.443)	(1.312)	(0.912)	(1.229)	(1.015)	(1.824)	(1.831)	(1.446)
	3	0.041	0.079	0.081	0.115	0.083	0.071	0.144	0.222	0.217	0.350	0.309	0.254
		(1.502)	(1.521)	(1.188)	(1.357)	(0.896)	(0.690)	(1.261)	(1.671)	(1.361)	(2.064)	(1.954)	(1.638)
	4	0.039	0.107	0.120	0.131	0.157	0.129	0.127	0.175	0.204	0.366	0.327	0.262
		(0.907)	(1.388)	(1.163)	(1.115)	(1.322)	(0.988)	(0.869)	(1.117)	(1.168)	(1.872)	(1.942)	(1.674)
	Hi FF3 R^2	0.100	0.131	0.165	0.196	0.123	0.152	0.195	0.263	0.284	0.322	0.222	0.213
		(1.291)	(1.096)	(1.175)	(1.232)	(0.698)	(0.813)	(0.942)	(1.258)	(1.310)	(1.481)	(1.285)	(1.259)
	Avg	0.036	0.064	0.075	0.098	0.082	0.083	0.107	0.155	0.164	0.298	0.262	0.202
		(1.272)	(1.338)	(1.244)	(1.346)	(1.053)	(0.957)	(1.072)	(1.404)	(1.307)	(1.944)	(1.990)	(1.652)
Act - Gho	Lo FF3 R^2	0.943	0.852	0.808	0.622	0.537	0.304	0.615	0.711	0.427	0.354	-0.589	-0.563
		(3.782)	(3.348)	(3.344)	(2.363)	(2.056)	(1.202)	(2.387)	(2.763)	(1.666)	(1.421)	(-3.073)	(-2.827)
	2	1.141	0.893	0.538	0.669	0.484	0.190	0.648	0.664	0.704	0.248	-0.893	-0.866
		(4.885)	(3.742)	(2.303)	(2.661)	(1.970)	(0.812)	(2.801)	(2.798)	(2.865)	(1.545)	(-4.599)	(-4.364)
	3	0.915	0.720	0.519	0.607	0.510	0.427	0.533	0.574	0.388	0.228	-0.686	-0.691
		(4.432)	(3.481)	(2.507)	(3.045)	(2.531)	(2.414)	(2.940)	(3.349)	(2.313)	(1.450)	(-4.009)	(-3.895)
	4	0.803	0.625	0.725	0.631	0.438	0.558	0.608	0.594	0.246	0.327	-0.476	-0.429
		(4.467)	(3.605)	(4.069)	(3.832)	(2.520)	(3.442)	(3.818)	(3.619)	(1.446)	(2.702)	(-3.043)	(-2.781)
	Hi FF3 R^2	0.662	0.663	0.500	0.286	0.614	0.391	0.409	0.303	0.265	0.366	-0.295	-0.313
		(4.159)	(4.128)	(3.181)	(1.731)	(4.062)	(2.672)	(2.894)	(2.246)	(2.041)	(2.943)	(-1.974)	(-2.158)
	Avg	0.893	0.750	0.618	0.563	0.517	0.374	0.563	0.569	0.406	0.305	-0.588	-0.572
		(5.084)	(4.391)	(3.787)	(3.418)	(3.061)	(2.450)	(3.654)	(3.779)	(2.812)	(2.700)	(-5.296)	(-5.469)

Table 28: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *FF5* R^2. We control for the effects of FF5 R^2. At the end of each month, we sort all the actual stock's $\mathsf{R}^2{}_{\mathrm{FF5},i,t-1}$ value into 5-tiles, which is obtained from regressing the past daily returns onto the Fama-French 5-factors according to the procedure of Section 5. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo'$, 2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo FF5 R^2	0.961	0.774	0.636	0.540	0.619	0.255	0.635	0.698	0.588	0.254	-0.707	-0.822
		(3.997)	(3.229)	(2.653)	(1.973)	(2.559)	(0.942)	(2.574)	(2.674)	(2.251)	(0.927)	(-3.341)	(-3.724)
	2	1.255	0.888	0.713	0.744	0.621	0.300	0.694	0.806	0.740	0.552	-0.703	-0.777
		(5.568)	(3.393)	(3.105)	(2.801)	(2.468)	(1.211)	(2.777)	(2.998)	(2.662)	(2.235)	(-3.735)	(-4.302)
	3	1.024	0.699	0.615	0.802	0.659	0.715	0.768	0.757	0.580	0.845	-0.179	-0.264
		(5.074)	(3.210)	(2.538)	(3.445)	(2.842)	(3.224)	(3.173)	(3.115)	(2.383)	(3.473)	(-1.086)	(-1.421)
	4	0.820	0.780	0.828	0.789	0.566	0.673	0.700	0.765	0.474	0.683	-0.137	-0.145
		(4.381)	(3.856)	(3.979)	(3.879)	(2.463)	(2.926)	(3.283)	(3.213)	(1.883)	(2.983)	(-0.773)	(-0.778)
	Hi FF5 R^2	0.722	0.790	0.631	0.574	0.730	0.540	0.573	0.592	0.538	0.689	-0.033	-0.066
		(4.034)	(4.056)	(2.735)	(2.493)	(3.101)	(2.259)	(2.145)	(2.447)	(2.102)	(2.793)	(-0.199)	(-0.383)
	Avg	0.956	0.786	0.684	0.690	0.639	0.496	0.674	0.724	0.584	0.605	-0.352	-0.415
		(5.275)	(4.024)	(3.414)	(3.236)	(3.024)	(2.357)	(3.114)	(3.265)	(2.571)	(2.819)	(-3.359)	(-3.977)
Ghost	Lo FF5 R^2	-0.001	-0.001	-0.000	-0.001	-0.001	0.002	0.008	0.024	0.018	0.156	0.156	0.071
		(-0.241)	(-0.394)	(-0.091)	(-0.443)	(-0.290)	(0.637)	(1.225)	(1.781)	(0.669)	(1.328)	(1.335)	(0.673)
	2	-0.001	0.001	0.008	0.059	0.043	0.076	0.069	0.066	0.105	0.228	0.229	0.151
		(-0.324)	(0.163)	(0.733)	(1.695)	(1.308)	(1.579)	(1.088)	(0.843)	(1.173)	(1.519)	(1.527)	(1.122)
	3	0.038	0.061	0.086	0.119	0.074	0.107	0.101	0.215	0.174	0.351	0.313	0.238
		(1.367)	(1.156)	(1.164)	(1.365)	(0.784)	(0.981)	(0.897)	(1.643)	(1.088)	(2.293)	(2.251)	(1.857)
	4	0.046	0.106	0.139	0.135	0.160	0.136	0.168	0.201	0.236	0.373	0.327	0.267
		(1.095)	(1.319)	(1.266)	(1.193)	(1.261)	(0.997)	(1.170)	(1.263)	(1.336)	(2.006)	(2.036)	(1.742)
	Hi FF5 R^2	0.092	0.134	0.152	0.197	0.132	0.137	0.206	0.278	0.281	0.337	0.245	0.234
		(1.261)	(1.126)	(1.061)	(1.230)	(0.760)	(0.709)	(0.969)	(1.369)	(1.276)	(1.558)	(1.401)	(1.364)
	Avg	0.035	0.060	0.077	0.102	0.081	0.091	0.110	0.157	0.163	0.289	0.254	0.192
		(1.279)	(1.246)	(1.216)	(1.383)	(1.016)	(1.011)	(1.100)	(1.437)	(1.300)	(1.944)	(1.983)	(1.641)
Act - Gho	Lo FF5 R^2	0.981	0.788	0.643	0.545	0.623	0.258	0.632	0.679	0.573	0.100	-0.881	-0.908
		(4.089)	(3.291)	(2.685)	(1.993)	(2.579)	(0.954)	(2.580)	(2.654)	(2.315)	(0.458)	(-4.535)	(-4.427)
	2	1.268	0.895	0.713	0.695	0.582	0.223	0.623	0.744	0.637	0.325	-0.942	-0.939
		(5.631)	(3.428)	(3.147)	(2.759)	(2.393)	(0.950)	(2.680)	(3.107)	(2.768)	(1.759)	(-4.953)	(-5.050)
	3	0.992	0.643	0.531	0.686	0.589	0.607	0.665	0.542	0.409	0.496	-0.496	-0.503
		(5.033)	(3.128)	(2.425)	(3.272)	(2.933)	(3.289)	(3.475)	(2.918)	(2.459)	(3.128)	(-3.029)	(-3.020)
	4	0.778	0.677	0.690	0.655	0.406	0.540	0.534	0.566	0.240	0.317	-0.461	-0.410
		(4.271)	(3.735)	(3.989)	(4.119)	(2.267)	(3.315)	(3.561)	(3.420)	(1.397)	(2.578)	(-2.888)	(-2.624)
	Hi FF5 R^2	0.632	0.656	0.481	0.377	0.599	0.403	0.369	0.314	0.258	0.357	-0.275	-0.298
		(3.984)	(4.119)	(2.981)	(2.501)	(3.876)	(2.869)	(2.527)	(2.413)	(1.965)	(2.887)	(-1.840)	(-2.048)
	Avg	0.930	0.732	0.612	0.592	0.560	0.406	0.565	0.569	0.423	0.319	-0.611	-0.612
		(5.395)	(4.232)	(3.699)	(3.508)	(3.328)	(2.612)	(3.696)	(3.697)	(2.999)	(2.785)	(-5.749)	(-5.998)

Table 29: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *market capitalization.* We control for the effects of market capitalization. At the end of each month, we sort all the actual stocks' market capitalization into 5-tiles using NYSE breakpoints. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3* α
Actual	Lo MktCap	1.543	1.026	0.774	0.873	0.522	0.624	0.735	0.968	1.004	1.033	-0.510	-0.439
		(4.255)	(2.853)	(2.159)	(2.324)	(1.601)	(1.762)	(2.008)	(2.379)	(2.695)	(2.315)	(-2.426)	(-2.034)
	2	1.355	0.736	0.859	0.706	0.442	0.490	0.749	0.780	0.758	0.746	-0.609	-0.585
		(4.459)	(2.560)	(2.999)	(2.435)	(1.404)	(1.521)	(2.260)	(2.156)	(2.023)	(1.929)	(-2.661)	(-2.611)
	3	1.215	1.089	0.782	0.675	0.683	0.786	0.797	0.784	0.762	0.836	-0.379	-0.353
		(4.669)	(4.131)	(2.772)	(2.369)	(2.336)	(2.670)	(2.614)	(2.579)	(2.343)	(2.554)	(-1.695)	(-1.511)
	4	1.050	1.014	0.949	0.703	0.861	0.799	0.939	0.883	0.794	0.791	-0.259	-0.210
		(4.197)	(4.225)	(3.946)	(2.984)	(3.616)	(3.100)	(3.601)	(3.160)	(2.724)	(2.665)	(-1.415)	(-1.101)
	Hi MktCap	0.761	0.697	0.679	0.736	0.545	0.593	0.616	0.569	0.612	0.719	-0.042	-0.084
		(4.572)	(4.225)	(3.863)	(3.959)	(2.760)	(2.704)	(2.824)	(2.258)	(2.618)	(3.085)	(-0.271)	(-0.511)
	Avg	1.185	0.912	0.809	0.738	0.611	0.659	0.767	0.797	0.786	0.825	-0.360	-0.334
		(4.794)	(3.790)	(3.242)	(2.910)	(2.376)	(2.438)	(2.768)	(2.663)	(2.631)	(2.647)	(-2.325)	(-2.088)
Ghost	Lo MktCap	-0.000	-0.000	-0.001	0.002	0.011	0.028	0.056	0.109	0.210	0.388	0.388	0.231
		(-0.075)	(-0.054)	(-0.433)	(0.809)	(1.626)	(1.680)	(1.660)	(1.726)	(1.716)	(1.543)	(1.545)	(1.021)
	2	-0.000	0.001	0.008	0.018	0.068	0.124	0.159	0.206	0.244	0.424	0.425	0.287
		(-0.211)	(0.255)	(0.890)	(0.732)	(1.322)	(1.454)	(1.225)	(1.226)	(1.190)	(1.475)	(1.478)	(1.102)
	3	0.017	0.074	0.105	0.130	0.130	0.175	0.203	0.230	0.269	0.418	0.401	0.302
		(0.879)	(1.013)	(0.968)	(1.003)	(0.926)	(1.107)	(1.164)	(1.156)	(1.158)	(1.501)	(1.470)	(1.197)
	4	0.035	0.087	0.118	0.139	0.198	0.232	0.259	0.264	0.305	0.421	0.386	0.336
		(0.790)	(0.962)	(1.072)	(1.080)	(1.329)	(1.374)	(1.385)	(1.240)	(1.262)	(1.550)	(1.523)	(1.413)
	Hi MktCap	0.037	0.076	0.144	0.150	0.144	0.141	0.151	0.250	0.277	0.367	0.330	0.293
		(1.179)	(1.212)	(1.507)	(1.325)	(1.099)	(0.939)	(0.850)	(1.319)	(1.381)	(1.777)	(1.716)	(1.567)
	Avg	0.018	0.047	0.075	0.088	0.110	0.140	0.165	0.212	0.261	0.404	0.386	0.290
		(0.974)	(1.080)	(1.221)	(1.169)	(1.223)	(1.277)	(1.256)	(1.357)	(1.384)	(1.634)	(1.630)	(1.340)
Act - Gho	Lo MktCap	1.548	1.031	0.780	0.874	0.515	0.601	0.684	0.863	0.801	0.663	-0.885	-0.662
		(4.272)	(2.869)	(2.175)	(2.330)	(1.587)	(1.720)	(1.945)	(2.307)	(2.662)	(2.530)	(-4.305)	(-3.727)
	2	1.361	0.739	0.854	0.690	0.377	0.371	0.592	0.576	0.517	0.323	-1.037	-0.877
		(4.479)	(2.573)	(2.999)	(2.449)	(1.275)	(1.301)	(2.182)	(2.115)	(2.105)	(1.698)	(-4.766)	(-4.401)
	3	1.205	1.019	0.680	0.547	0.554	0.614	0.597	0.556	0.490	0.419	-0.786	-0.661
		(4.671)	(4.181)	(2.824)	(2.340)	(2.398)	(2.831)	(2.824)	(2.817)	(2.695)	(2.874)	(-3.636)	(-3.284)
	4	1.021	0.930	0.834	0.565	0.666	0.569	0.681	0.622	0.492	0.376	-0.645	-0.546
		(4.271)	(4.392)	(4.325)	(3.057)	(3.922)	(3.347)	(4.022)	(3.726)	(3.215)	(2.875)	(-3.023)	(-2.851)
	Hi MktCap	0.729	0.622	0.536	0.588	0.402	0.453	0.466	0.321	0.336	0.356	-0.373	-0.379
		(4.546)	(4.114)	(3.607)	(3.957)	(2.847)	(3.087)	(3.555)	(2.295)	(2.711)	(3.082)	(-2.380)	(-2.425)
	Avg	1.173	0.868	0.737	0.653	0.503	0.521	0.604	0.588	0.527	0.427	-0.745	-0.625
		(4.853)	(3.895)	(3.403)	(3.101)	(2.477)	(2.592)	(3.169)	(3.080)	(3.304)	(3.563)	(-4.528)	(-4.429)

Table 30: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *book-to-market*. We control for value effects. At the end of each month, we sort all the actual stocks' book-to-market ratios into 5-tiles using NYSE breakpoints. We follow the standard Fama and French (1992, 1993) procedure to define firm i's book-to-market ratio at month $t-1$. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo B/M	1.005	0.588	0.452	0.310	0.092	0.301	0.582	0.339	0.625	0.484	-0.521	-0.470
		(4.008)	(2.595)	(1.936)	(1.324)	(0.393)	(1.208)	(2.281)	(1.185)	(2.316)	(1.591)	(-2.350)	(-2.138)
	2	0.945	0.638	0.696	0.625	0.886	0.804	0.692	0.653	0.646	0.660	-0.285	-0.289
		(4.375)	(2.954)	(3.226)	(3.051)	(3.878)	(3.591)	(3.148)	(2.983)	(2.692)	(2.946)	(-1.627)	(-1.446)
	3	0.749	1.095	0.770	0.834	0.662	0.787	0.863	1.017	0.777	0.616	-0.133	-0.145
		(3.594)	(5.061)	(3.306)	(3.749)	(3.173)	(3.791)	(3.906)	(4.582)	(3.360)	(2.682)	(-0.825)	(-0.867)
	4	0.886	0.870	0.799	0.816	0.772	0.834	1.012	0.608	0.707	0.811	-0.075	-0.076
		(4.096)	(4.005)	(3.765)	(3.225)	(3.476)	(3.796)	(4.693)	(2.498)	(2.817)	(3.300)	(-0.456)	(-0.443)
	Hi B/M	1.113	0.771	0.984	0.768	0.781	1.030	1.151	0.855	1.323	0.968	-0.145	-0.085
		(3.949)	(2.790)	(3.766)	(2.781)	(2.898)	(3.864)	(4.144)	(3.228)	(5.524)	(4.018)	(-0.718)	(-0.402)
	Avg	0.940	0.792	0.740	0.671	0.638	0.751	0.860	0.694	0.816	0.708	-0.232	-0.213
		(4.603)	(4.058)	(3.804)	(3.378)	(3.246)	(3.783)	(4.190)	(3.280)	(3.769)	(3.178)	(-1.980)	(-1.773)
Ghost	Lo B/M	-0.000	0.034	0.044	0.006	-0.011	0.099	0.103	0.073	0.204	0.143	0.143	0.103
		(-0.001)	(1.625)	(0.908)	(0.093)	(-0.130)	(0.881)	(0.691)	(0.417)	(1.010)	(0.601)	(0.602)	(0.456)
	2	-0.001	0.031	0.057	0.082	0.109	0.124	0.122	0.223	0.238	0.352	0.353	0.284
		(-0.591)	(1.326)	(1.265)	(1.141)	(1.167)	(1.170)	(0.956)	(1.474)	(1.390)	(1.823)	(1.833)	(1.542)
	3	-0.001	0.012	0.084	0.063	0.091	0.140	0.161	0.246	0.234	0.345	0.346	0.258
		(-0.409)	(0.933)	(1.791)	(0.956)	(1.028)	(1.301)	(1.269)	(1.502)	(1.349)	(1.840)	(1.848)	(1.424)
	4	-0.000	-0.001	0.026	0.075	0.080	0.102	0.149	0.181	0.209	0.363	0.364	0.242
		(-0.211)	(-0.424)	(1.328)	(1.457)	(0.964)	(0.932)	(1.140)	(1.225)	(1.195)	(1.724)	(1.729)	(1.168)
	Hi B/M	0.000	-0.002	0.006	0.021	0.055	0.124	0.100	0.161	0.346	0.642	0.642	0.517
		(0.022)	(-0.786)	(1.594)	(1.685)	(1.290)	(1.404)	(0.851)	(1.113)	(2.056)	(3.170)	(3.172)	(2.811)
	Avg	-0.001	0.015	0.044	0.050	0.065	0.118	0.127	0.177	0.246	0.369	0.370	0.281
		(-0.246)	(1.374)	(1.427)	(1.017)	(0.878)	(1.194)	(1.039)	(1.198)	(1.441)	(1.873)	(1.878)	(1.506)
Act - Gho	Lo B/M	1.008	0.556	0.411	0.305	0.105	0.203	0.480	0.269	0.422	0.344	-0.664	-0.574
		(4.021)	(2.538)	(1.856)	(1.471)	(0.544)	(0.997)	(2.440)	(1.350)	(2.521)	(2.273)	(-2.724)	(-2.542)
	2	0.954	0.611	0.640	0.544	0.778	0.681	0.569	0.435	0.410	0.312	-0.642	-0.577
		(4.423)	(2.880)	(3.096)	(2.908)	(4.028)	(3.689)	(3.388)	(2.660)	(2.738)	(2.440)	(-2.819)	(-2.434)
	3	0.758	1.086	0.688	0.774	0.573	0.645	0.703	0.774	0.543	0.273	-0.485	-0.410
		(3.644)	(5.080)	(3.077)	(3.770)	(3.203)	(3.955)	(4.366)	(5.216)	(3.702)	(2.100)	(-2.439)	(-2.084)
	4	0.894	0.875	0.776	0.737	0.693	0.733	0.864	0.427	0.497	0.447	-0.447	-0.325
		(4.141)	(4.033)	(3.718)	(3.048)	(3.557)	(3.968)	(5.474)	(2.277)	(2.983)	(3.704)	(-2.157)	(-1.693)
	Hi B/M	1.119	0.777	0.980	0.748	0.727	0.910	1.049	0.695	0.974	0.328	-0.791	-0.606
		(3.972)	(2.810)	(3.762)	(2.751)	(2.821)	(3.793)	(4.653)	(3.278)	(5.898)	(2.196)	(-2.721)	(-2.224)
	Avg	0.947	0.781	0.699	0.622	0.575	0.635	0.733	0.520	0.569	0.341	-0.606	-0.498
		(4.643)	(4.067)	(3.766)	(3.475)	(3.516)	(4.118)	(5.088)	(3.689)	(4.584)	(3.201)	(-3.134)	(-2.823)

Table 31: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *dollar volume liquidity.* We control for the effects of dollar volume liquidity. At the end of each month, we sort all the actual stocks' dollar volume liquidity into 5-tiles. Dollar volume liquidity of stock i at month $t-1$ is defined as $\text{VOLD}_{i,t-1} :=$ (Trading volume of stock i traded on the last day of month $t-1$) $\times$ (Closing price of stock i on the last day of month $t-1$) $\div$ $1,000,000$. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo VOLD	1.241	1.020	1.007	1.182	0.732	0.716	0.979	0.893	1.027	1.057	-0.184	-0.089
		(4.517)	(3.714)	(3.920)	(4.192)	(2.650)	(2.563)	(3.663)	(3.162)	(3.747)	(3.927)	(-1.215)	(-0.554)
	2	1.231	0.919	0.903	0.882	0.644	0.799	1.073	0.915	1.130	0.906	-0.325	-0.264
		(5.095)	(3.474)	(3.425)	(3.524)	(2.324)	(2.921)	(3.888)	(3.278)	(4.319)	(3.755)	(-2.024)	(-1.595)
	3	1.113	1.006	0.888	0.865	0.747	0.738	0.802	0.933	0.938	0.883	-0.230	-0.258
		(4.628)	(4.284)	(3.567)	(3.211)	(3.124)	(3.007)	(3.203)	(3.715)	(3.721)	(3.412)	(-1.240)	(-1.348)
	4	0.966	0.961	0.798	0.681	0.786	0.776	0.818	0.805	0.754	0.947	-0.019	-0.017
		(4.297)	(4.562)	(3.834)	(3.107)	(3.639)	(3.419)	(3.386)	(3.332)	(2.977)	(3.808)	(-0.121)	(-0.106)
	Hi VOLD	0.800	0.654	0.603	0.634	0.504	0.584	0.644	0.513	0.567	0.708	-0.092	-0.140
		(4.771)	(4.023)	(3.259)	(3.201)	(2.312)	(2.677)	(2.831)	(2.009)	(2.343)	(3.009)	(-0.567)	(-0.830)
	Avg	1.070	0.912	0.840	0.849	0.683	0.723	0.863	0.812	0.883	0.900	-0.170	-0.154
		(5.140)	(4.318)	(3.980)	(3.791)	(3.016)	(3.159)	(3.700)	(3.359)	(3.768)	(3.908)	(-1.547)	(-1.347)
Ghost	Lo VOLD	-0.000	-0.001	-0.001	-0.001	0.004	0.007	0.031	0.039	0.093	0.296	0.297	0.237
		(-0.199)	(-0.560)	(-0.401)	(-0.602)	(1.053)	(0.937)	(1.704)	(1.230)	(1.455)	(2.170)	(2.175)	(1.909)
	2	-0.000	0.005	0.011	0.037	0.064	0.129	0.140	0.137	0.192	0.333	0.333	0.251
		(-0.199)	(0.752)	(0.743)	(1.006)	(1.190)	(1.549)	(1.292)	(0.999)	(1.198)	(1.696)	(1.702)	(1.398)
	3	0.007	0.079	0.101	0.092	0.125	0.156	0.200	0.224	0.258	0.350	0.343	0.264
		(0.450)	(1.032)	(1.023)	(0.801)	(1.013)	(1.050)	(1.241)	(1.332)	(1.382)	(1.586)	(1.599)	(1.355)
	4	0.037	0.072	0.122	0.133	0.179	0.200	0.250	0.252	0.225	0.429	0.391	0.333
		(0.822)	(0.936)	(1.203)	(1.115)	(1.371)	(1.289)	(1.464)	(1.357)	(1.084)	(1.926)	(1.917)	(1.785)
	Hi VOLD	0.035	0.085	0.141	0.103	0.142	0.099	0.178	0.270	0.251	0.358	0.323	0.287
		(1.115)	(1.242)	(1.412)	(0.844)	(0.991)	(0.602)	(0.992)	(1.355)	(1.215)	(1.738)	(1.688)	(1.532)
	Avg	0.016	0.048	0.075	0.073	0.103	0.118	0.160	0.185	0.204	0.353	0.337	0.274
		(0.863)	(1.113)	(1.271)	(0.986)	(1.196)	(1.121)	(1.319)	(1.340)	(1.309)	(1.876)	(1.892)	(1.688)
Act - Gho	Lo VOLD	1.246	1.025	1.011	1.186	0.731	0.712	0.952	0.857	0.935	0.766	-0.480	-0.326
		(4.540)	(3.735)	(3.936)	(4.205)	(2.653)	(2.568)	(3.651)	(3.209)	(3.885)	(4.030)	(-2.681)	(-1.852)
	2	1.236	0.918	0.895	0.846	0.582	0.671	0.935	0.779	0.939	0.587	-0.649	-0.505
		(5.116)	(3.477)	(3.444)	(3.534)	(2.269)	(2.804)	(4.072)	(3.612)	(4.973)	(3.737)	(-2.935)	(-2.564)
	3	1.111	0.931	0.787	0.775	0.623	0.582	0.604	0.710	0.678	0.536	-0.575	-0.524
		(4.676)	(4.294)	(3.575)	(3.494)	(3.325)	(3.194)	(3.523)	(4.127)	(4.292)	(3.633)	(-2.633)	(-2.524)
	4	0.935	0.891	0.678	0.547	0.610	0.579	0.569	0.554	0.535	0.524	-0.411	-0.350
		(4.379)	(4.700)	(4.126)	(3.253)	(3.960)	(3.895)	(3.599)	(3.701)	(3.575)	(3.986)	(-2.015)	(-1.825)
	Hi VOLD	0.769	0.570	0.464	0.534	0.363	0.485	0.466	0.243	0.318	0.354	-0.415	-0.428
		(4.776)	(3.892)	(3.043)	(3.599)	(2.425)	(3.423)	(3.432)	(1.708)	(2.552)	(3.030)	(-2.654)	(-2.710)
	Avg	1.059	0.867	0.767	0.778	0.582	0.606	0.705	0.629	0.681	0.553	-0.506	-0.427
		(5.227)	(4.469)	(4.229)	(4.281)	(3.331)	(3.671)	(4.422)	(4.074)	(5.149)	(4.850)	(-3.130)	(-2.906)

Table 32: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and *Amihud's illiquidity.* We control for the effects of Amihud's measure of illiquidity. At the end of each month, we sort all the actual stocks' Amihud's illiquidity into 5-tiles. Following Amihud (2002), the illiquidity measure for stock i at month $t-1$ is defined as $\text{AmihudIlliq}_{i,t-1} := 1/D_{i,t-1} \sum_{d=1}^{D_{i,t-1}} \left|R_{i,d}\right| / \text{VOLD}_{i,d}$, where $D_{i,t-1}$ is the total number of trading days of stock i in the past twelve months leading up to month $t-1$, and VOLD is dollar volume illiquidity, defined as $\text{VOLD}_{i,t-1} :=$ (Trading volume of stock i traded on the last day of month $t-1$) $\times$ (Closing price of stock i on the last day of month $t-1$) $\div$ $1,000,000$. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo AmihudIlliq	0.793	0.661	0.680	0.606	0.586	0.605	0.620	0.507	0.585	0.719	-0.074	-0.132
		(4.834)	(4.093)	(3.958)	(3.029)	(2.996)	(2.771)	(2.808)	(1.953)	(2.442)	(3.079)	(-0.459)	(-0.781)
	2	0.887	0.947	0.789	0.747	0.711	0.807	0.764	0.768	0.739	0.832	-0.055	-0.005
		(3.902)	(4.213)	(3.495)	(3.310)	(3.020)	(3.292)	(2.953)	(2.850)	(2.581)	(2.966)	(-0.330)	(-0.031)
	3	1.221	1.126	0.853	0.801	0.620	0.685	0.780	0.806	0.795	0.807	-0.414	-0.444
		(4.675)	(4.325)	(3.099)	(2.781)	(2.423)	(2.572)	(2.887)	(2.888)	(2.819)	(2.914)	(-2.074)	(-2.101)
	4	1.387	1.025	0.922	0.816	0.785	0.628	1.002	0.918	1.145	0.958	-0.429	-0.383
		(4.790)	(3.780)	(3.550)	(3.039)	(2.687)	(2.226)	(3.413)	(3.108)	(3.682)	(3.200)	(-2.040)	(-1.726)
	Hi AmihudIlliq	1.087	0.990	0.840	0.915	0.643	0.635	0.873	0.759	0.782	1.018	-0.069	-0.001
		(3.065)	(3.117)	(3.072)	(2.999)	(2.105)	(2.023)	(2.812)	(2.309)	(2.563)	(3.228)	(-0.320)	(-0.004)
	Avg	1.075	0.950	0.817	0.777	0.669	0.672	0.808	0.752	0.809	0.867	-0.208	-0.193
		(4.578)	(4.205)	(3.689)	(3.280)	(2.804)	(2.774)	(3.245)	(2.848)	(3.093)	(3.341)	(-1.596)	(-1.427)
Ghost	Lo AmihudIlliq	0.036	0.074	0.155	0.131	0.152	0.113	0.171	0.249	0.260	0.361	0.325	0.287
		(1.154)	(1.133)	(1.641)	(1.112)	(1.168)	(0.711)	(0.967)	(1.263)	(1.280)	(1.742)	(1.684)	(1.523)
	2	0.043	0.084	0.117	0.167	0.191	0.223	0.261	0.240	0.257	0.421	0.378	0.331
		(0.885)	(0.990)	(1.078)	(1.305)	(1.276)	(1.281)	(1.383)	(1.144)	(1.100)	(1.644)	(1.601)	(1.514)
	3	0.027	0.108	0.113	0.110	0.142	0.162	0.228	0.231	0.223	0.357	0.330	0.264
		(0.944)	(1.147)	(0.990)	(0.859)	(0.951)	(0.989)	(1.281)	(1.215)	(1.041)	(1.430)	(1.380)	(1.197)
	4	-0.001	0.004	0.011	0.029	0.068	0.113	0.138	0.193	0.226	0.334	0.335	0.209
		(-0.291)	(0.736)	(0.791)	(1.010)	(1.266)	(1.446)	(1.182)	(1.308)	(1.211)	(1.399)	(1.404)	(0.977)
	Hi AmihudIlliq	-0.000	-0.001	-0.001	-0.002	0.006	0.015	0.037	0.040	0.109	0.397	0.397	0.320
		(-0.004)	(-0.497)	(-0.374)	(-0.966)	(1.112)	(1.328)	(1.462)	(0.993)	(1.410)	(2.319)	(2.321)	(2.032)
	Avg	0.021	0.054	0.079	0.087	0.112	0.125	0.167	0.191	0.215	0.374	0.353	0.282
		(0.990)	(1.138)	(1.275)	(1.143)	(1.218)	(1.132)	(1.292)	(1.285)	(1.250)	(1.742)	(1.736)	(1.526)
Act - Gho	Lo AmihudIlliq	0.761	0.589	0.526	0.478	0.434	0.492	0.448	0.259	0.326	0.362	-0.398	-0.420
		(4.824)	(4.007)	(3.600)	(3.146)	(3.096)	(3.525)	(3.324)	(1.818)	(2.630)	(3.148)	(-2.617)	(-2.699)
	2	0.850	0.866	0.674	0.579	0.523	0.587	0.505	0.531	0.489	0.414	-0.435	-0.338
		(3.970)	(4.342)	(3.832)	(3.391)	(3.388)	(3.671)	(3.139)	(3.469)	(3.273)	(3.160)	(-2.165)	(-1.889)
	3	1.200	1.022	0.741	0.693	0.480	0.524	0.555	0.576	0.571	0.454	-0.746	-0.710
		(4.687)	(4.410)	(3.127)	(2.993)	(2.428)	(2.636)	(3.031)	(3.140)	(3.373)	(2.884)	(-3.068)	(-2.918)
	4	1.394	1.026	0.914	0.790	0.719	0.516	0.868	0.726	0.921	0.635	-0.759	-0.586
		(4.818)	(3.797)	(3.560)	(3.040)	(2.653)	(2.086)	(3.585)	(3.181)	(4.270)	(3.509)	(-3.065)	(-2.541)
	Hi AmihudIlliq	1.093	0.996	0.844	0.920	0.640	0.623	0.842	0.723	0.676	0.636	-0.457	-0.314
		(3.083)	(3.138)	(3.090)	(3.015)	(2.099)	(2.001)	(2.797)	(2.330)	(2.523)	(2.899)	(-1.842)	(-1.284)
	Avg	1.059	0.900	0.740	0.692	0.559	0.548	0.644	0.563	0.597	0.500	-0.559	-0.473
		(4.613)	(4.339)	(3.886)	(3.591)	(3.076)	(3.123)	(3.846)	(3.402)	(4.203)	(4.174)	(-3.134)	(-2.905)

Table 33: Value-weight portfolios formed by dependent bivariate sort of elastic-net R^2 and *short-term reversal.* We control for short-term reversal effects. At the end of each month, we sort all the actual stocks' short-term-reversal into 5-tiles. Following Jegadeesh (1990) and Lehmann (1990), short-term reversal is defined as $\text{STR}_{i,t-1} = 100 \times R_{i,t-1}$, where $R_{i,t-1}$ is the month $t-1$ return of actual stock i. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \ldots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo STR	1.419	1.283	1.000	1.016	0.740	0.784	0.996	0.916	0.655	0.418	-1.001	-0.927
		(4.212)	(4.055)	(2.966)	(3.238)	(2.541)	(2.491)	(3.070)	(3.018)	(2.081)	(1.179)	(-3.330)	(-3.089)
	2	0.987	0.960	1.096	0.545	0.869	0.839	1.035	0.781	0.919	0.898	-0.089	0.002
		(3.913)	(4.336)	(4.857)	(2.227)	(3.971)	(3.616)	(4.651)	(3.353)	(3.722)	(3.467)	(-0.415)	(0.008)
	3	1.027	0.769	0.617	0.650	0.665	0.796	0.758	0.824	0.816	0.733	-0.294	-0.155
		(4.868)	(3.411)	(2.969)	(3.779)	(3.515)	(4.130)	(3.809)	(3.804)	(3.553)	(3.276)	(-1.589)	(-0.839)
	4	0.902	0.813	0.340	0.668	0.586	0.728	0.486	0.496	0.537	0.766	-0.135	-0.117
		(3.836)	(4.006)	(1.678)	(3.558)	(2.600)	(3.621)	(2.363)	(2.261)	(2.310)	(3.067)	(-0.645)	(-0.505)
	Hi STR	0.915	0.641	0.277	0.224	0.345	0.209	0.540	0.464	0.334	0.501	-0.414	-0.478
		(3.430)	(2.194)	(1.044)	(0.935)	(1.395)	(0.821)	(2.226)	(1.736)	(1.231)	(1.782)	(-1.695)	(-1.905)
	Avg	1.050	0.893	0.666	0.621	0.641	0.671	0.763	0.696	0.652	0.663	-0.387	-0.335
		(4.952)	(4.246)	(3.256)	(3.153)	(3.273)	(3.428)	(3.661)	(3.246)	(2.806)	(2.664)	(-2.446)	(-2.162)
Ghost	Lo STR	-0.001	-0.011	-0.003	0.040	0.004	0.109	0.092	0.149	0.137	0.189	0.190	0.139
		(-0.320)	(-1.113)	(-0.090)	(0.818)	(0.043)	(1.029)	(0.680)	(0.894)	(0.661)	(0.701)	(0.705)	(0.512)
	2	-0.001	0.017	0.022	0.038	0.068	0.069	0.168	0.183	0.199	0.320	0.321	0.276
		(-0.316)	(1.268)	(0.652)	(0.702)	(0.877)	(0.759)	(1.364)	(1.275)	(1.094)	(1.516)	(1.521)	(1.247)
	3	-0.001	0.019	0.040	0.051	0.086	0.154	0.158	0.210	0.215	0.383	0.384	0.345
		(-0.459)	(1.264)	(1.204)	(1.084)	(1.059)	(1.639)	(1.419)	(1.543)	(1.272)	(2.047)	(2.054)	(1.870)
	4	-0.001	0.031	0.076	0.087	0.133	0.161	0.187	0.178	0.265	0.414	0.415	0.311
		(-0.442)	(1.392)	(1.664)	(1.260)	(1.384)	(1.521)	(1.532)	(1.293)	(1.455)	(2.003)	(2.012)	(1.655)
	Hi STR	-0.001	0.020	0.031	0.087	0.106	0.085	0.189	0.190	0.280	0.259	0.261	0.112
		(-0.634)	(0.927)	(0.896)	(1.100)	(1.041)	(0.719)	(1.211)	(1.139)	(1.392)	(1.066)	(1.074)	(0.502)
	Avg	-0.001	0.015	0.033	0.061	0.079	0.116	0.159	0.182	0.219	0.313	0.314	0.237
		(-0.462)	(1.498)	(1.187)	(1.111)	(0.952)	(1.184)	(1.282)	(1.249)	(1.200)	(1.469)	(1.476)	(1.158)
Act - Gho	Lo STR	1.424	1.295	1.004	0.977	0.739	0.683	0.905	0.768	0.515	0.235	-1.188	-1.065
		(4.225)	(4.102)	(3.038)	(3.303)	(2.747)	(2.548)	(3.742)	(3.994)	(2.666)	(1.341)	(-3.515)	(-3.261)
	2	0.993	0.946	1.076	0.508	0.802	0.769	0.868	0.599	0.718	0.583	-0.409	-0.276
		(3.933)	(4.351)	(5.093)	(2.280)	(4.181)	(4.008)	(4.921)	(3.512)	(4.530)	(4.369)	(-1.715)	(-1.181)
	3	1.036	0.758	0.579	0.602	0.581	0.642	0.602	0.615	0.601	0.353	-0.683	-0.506
		(4.915)	(3.351)	(2.853)	(3.650)	(3.507)	(3.834)	(3.725)	(3.785)	(3.968)	(2.801)	(-3.282)	(-2.553)
	4	0.910	0.789	0.264	0.582	0.454	0.569	0.299	0.322	0.274	0.354	-0.556	-0.433
		(3.877)	(3.893)	(1.292)	(3.423)	(2.339)	(3.304)	(1.799)	(1.883)	(1.792)	(2.486)	(-2.222)	(-1.814)
	Hi STR	0.922	0.625	0.248	0.139	0.240	0.122	0.352	0.277	0.053	0.245	-0.678	-0.594
		(3.467)	(2.200)	(0.957)	(0.653)	(1.171)	(0.565)	(1.774)	(1.348)	(0.306)	(1.873)	(-2.395)	(-2.144)
	Avg	1.057	0.883	0.634	0.562	0.563	0.557	0.605	0.516	0.432	0.354	-0.703	-0.575
		(4.989)	(4.250)	(3.261)	(3.233)	(3.500)	(3.698)	(4.018)	(3.903)	(3.297)	(3.204)	(-3.396)	(-3.047)

Table 34: Value-weighted portfolios formed by dependent bivariate sort of elastic-net R^2 and momentum. We control for the effects of momentum. We define momentum at the end of month $t-1$ (i.e. the end of the estimation period) as the return of the stock during the 11-month period covering months $t-12$ through $t-2$; specifically, $\text{Mom}_{i,t-1} = 100 * \left[\prod_{m\in\{t-12:t-2\}}(R_{i,m}+1)-1\right]$, where $R_{i,m}$ is the return of stock i in month m. Then within each of these 5-tiles, we further the actual stocks' end of month elastic-net R^2 into same sized 10-tiles. The "Actual" group refers to the portfolio of actual stocks as per (10), the "Ghost" group refers to the portfolio of ghost stocks as per (11), while the "Act - Gho" group refers to the portfolio of long-short actuals against ghost stocks as per (12). The weighting scheme of the actual and ghost stocks are formed by according to Section 2.2. For each $k =$ 'Lo', $2, \dots,$ 'Hi', the row "Avg" reports the simple average across the five control variable portfolio bins. The column "H-L" is the "Hi" portfolio bin return, less the "Lo" portfolio bin return. The column "H-L FF3 α" is the Fama-French 3 factor alpha estimate of (13) for the "Hi" subtract "Lo" portfolio bin return. Portfolios are held for one month and returns in monthly percentage points are reported (e.g. 1.0 means 1%). The parentheses show robust Newey and West (1987) t-statistics with 6 lags. The time sample is from December 1975 to December 2017.

		Lo EN R^2	2	3	4	5	6	7	8	9	Hi EN R^2	*H-L*	*H-L FF3 α*
Actual	Lo Mom	-1.988	-2.541	-2.264	-2.690	-2.469	-2.338	-2.152	-2.343	-2.120	-2.285	-0.297	-0.223
		(-5.462)	(-7.260)	(-6.900)	(-7.694)	(-7.359)	(-7.290)	(-6.462)	(-7.131)	(-6.259)	(-6.256)	(-0.957)	(-0.722)
	2	-0.113	-0.077	-0.086	-0.299	-0.216	-0.139	-0.046	-0.268	-0.219	-0.363	-0.249	-0.178
		(-0.495)	(-0.326)	(-0.334)	(-1.439)	(-0.972)	(-0.584)	(-0.203)	(-1.096)	(-0.941)	(-1.444)	(-1.306)	(-0.943)
	3	0.839	0.568	0.605	0.427	0.590	0.689	0.711	0.621	0.577	0.626	-0.212	-0.249
		(4.295)	(2.793)	(2.679)	(2.276)	(2.957)	(3.383)	(3.470)	(2.959)	(2.651)	(2.888)	(-1.240)	(-1.241)
	4	1.503	1.485	1.427	1.482	1.370	1.282	1.426	1.341	1.322	1.318	-0.184	-0.168
		(6.699)	(6.622)	(7.341)	(7.300)	(7.114)	(5.893)	(6.810)	(6.154)	(5.851)	(5.632)	(-1.138)	(-1.032)
	Hi Mom	3.619	3.323	3.194	2.917	2.823	2.985	3.022	3.085	2.904	2.611	-1.008	-0.932
		(11.011)	(9.856)	(10.401)	(9.813)	(9.461)	(10.155)	(9.941)	(10.697)	(9.620)	(8.182)	(-3.849)	(-3.549)
	Avg	0.772	0.552	0.575	0.367	0.419	0.496	0.592	0.487	0.493	0.382	-0.390	-0.350
		(3.463)	(2.398)	(2.544)	(1.745)	(1.917)	(2.251)	(2.621)	(2.137)	(2.052)	(1.519)	(-2.755)	(-2.340)
Ghost	Lo Mom	-0.001	0.000	0.016	0.014	-0.009	0.025	0.080	-0.030	-0.074	-0.124	-0.123	-0.197
		(-0.312)	(0.150)	(1.011)	(0.404)	(-0.126)	(0.239)	(0.546)	(-0.160)	(-0.322)	(-0.458)	(-0.456)	(-0.765)
	2	-0.001	0.014	0.025	0.063	0.052	0.106	0.083	0.084	0.138	0.119	0.120	0.082
		(-0.309)	(1.307)	(0.888)	(1.366)	(0.661)	(1.014)	(0.638)	(0.590)	(0.780)	(0.596)	(0.601)	(0.415)
	3	-0.001	0.018	0.042	0.051	0.097	0.146	0.147	0.165	0.223	0.339	0.340	0.244
		(-0.469)	(1.248)	(1.078)	(1.026)	(1.316)	(1.393)	(1.376)	(1.123)	(1.274)	(1.778)	(1.784)	(1.325)
	4	-0.001	0.026	0.076	0.109	0.096	0.138	0.206	0.286	0.278	0.451	0.452	0.370
		(-0.410)	(1.425)	(1.625)	(1.555)	(1.167)	(1.361)	(1.770)	(1.890)	(1.567)	(2.211)	(2.218)	(1.889)
	Hi Mom	-0.000	0.017	0.036	0.072	0.130	0.149	0.232	0.226	0.430	0.683	0.683	0.577
		(-0.058)	(1.282)	(1.149)	(1.173)	(1.352)	(1.150)	(1.437)	(1.280)	(2.076)	(2.499)	(2.503)	(2.191)
	Avg	-0.001	0.015	0.039	0.062	0.073	0.113	0.150	0.146	0.199	0.294	0.294	0.215
		(-0.318)	(1.505)	(1.355)	(1.236)	(0.950)	(1.076)	(1.176)	(0.947)	(1.063)	(1.344)	(1.349)	(1.027)
Act - Gho	Lo Mom	-1.987	-2.541	-2.279	-2.704	-2.460	-2.362	-2.237	-2.314	-2.046	-2.161	-0.174	-0.026
		(-5.461)	(-7.267)	(-7.022)	(-8.072)	(-7.961)	(-8.675)	(-8.379)	(-9.625)	(-9.879)	(-11.094)	(-0.509)	(-0.077)
	2	-0.113	-0.091	-0.111	-0.362	-0.268	-0.245	-0.129	-0.351	-0.358	-0.482	-0.370	-0.259
		(-0.491)	(-0.387)	(-0.450)	(-1.864)	(-1.414)	(-1.213)	(-0.768)	(-1.966)	(-2.476)	(-4.184)	(-1.733)	(-1.290)
	3	0.840	0.550	0.563	0.376	0.493	0.542	0.564	0.457	0.354	0.287	-0.553	-0.493
		(4.301)	(2.720)	(2.604)	(2.132)	(2.789)	(3.304)	(3.520)	(3.181)	(2.604)	(2.277)	(-2.768)	(-2.298)
	4	1.504	1.459	1.351	1.373	1.274	1.144	1.220	1.055	1.045	0.866	-0.637	-0.539
		(6.701)	(6.658)	(7.223)	(7.632)	(7.702)	(6.483)	(7.243)	(6.300)	(7.546)	(6.732)	(-2.836)	(-2.516)
	Hi Mom	3.620	3.306	3.158	2.845	2.693	2.836	2.789	2.859	2.473	1.928	-1.692	-1.509
		(11.014)	(9.920)	(10.745)	(10.131)	(10.152)	(12.342)	(12.285)	(14.291)	(12.638)	(11.284)	(-4.775)	(-5.154)
	Avg	0.773	0.537	0.536	0.306	0.346	0.383	0.442	0.341	0.294	0.088	-0.685	-0.565
		(3.468)	(2.362)	(2.490)	(1.599)	(1.858)	(2.284)	(2.754)	(2.333)	(2.314)	(0.803)	(-3.251)	(-2.951)