Cognitive Errors As Canary Traps*

Alex Chinco†

November 18, 2021

Abstract

When otherwise intelligent investors fail to correct an error, a researcher learns something about what these investors did not know. The investors must not have known about anything which would have allowed them to spot their mistake. If they had, they would have stopped making it.

I show how a researcher can use this insight to identify how investors price assets. If X predicts returns, then any correlated predictor that investors know about but researchers have yet to discover represents a potential confound. I define a special kind of error, called a "cognitive error", which otherwise intelligent investors will only fail to correct if they are not aware of any such omitted variables. So when investors fail to correct a priced cognitive error about X, then a researcher can be sure that investors are pricing assets based on X. Cognitive errors are instruments for identifying how mostly rational investors price assets.

Keywords: Model Identification, Cognitive Errors, Market Efficiency

---

*I would like to thank Justin Birru, Brian Boyer, Svetlana Bryzgalova, Andrew Chen, James Dow, Arpit Gupta, Sam Hartzmark, Christian Heyerdahl-Larsen, Jiacui Li, Andy Neuhierl, Cameron Peng, Marco Sammon, Shri Santosh, Eric So, Michael Weber, and Jeff Wurgler as well as participants at the European Winter Finance Conference and Baruch for excellent comments and suggestions.

†Baruch College, Zicklin School of Business. alexchinco@gmail.com

Current Version: http://www.alexchinco.com/canary-traps.pdf

1 Introduction

Imagine you are a PhD student interested in the cross-section of returns.

A "value stock" has a price that seems low compared to some measure of fundamentals, such as its earnings or book value. The opposite is true for a "growth stock". We have long known that value stocks tend to have higher returns than growth stocks (Graham and Dodd, 1934). This is the "value premium".

Your job market paper documents that, in addition to having higher earnings-to-price and book-to-market ratios than growth stocks, value stocks also tend to have lower levels of another variable, X. You argue that the value premium exists because investors have been pricing assets based on X all along. Ratios of price to fundamentals are just a sideshow. Investors do not consider these things when computing prices.

You are not the first person to write a job market paper on the value premium. But you are hoping to be the last. Your goal is to definitively prove that X is the real reason for this well-known empirical regularity. You already have results showing that X subsumes all previous explanations. And your job market paper also verifies several novel predictions which, as far as you know, are only consistent with an X-based story.

But if outperforming previous explanations and making novel predictions were enough, the value premium would have been solved long ago. Every previous paper also did these things. That is how they got published. Now you are arguing that investors have been pricing assets based on X all along even though no previous researcher ever thought of the variable. So you have to consider the possibility that your X will be explained away by some future PhD student using a variable you yourself have never thought of.

How can you identify that investors are pricing assets based on X when you know that there will be more discoveries coming down the research pipeline? Is it even possible? This is a classic omitted-variables problem. The existing explanations for the value premium represent observed confounds. The explanations that future researchers will propose are omitted variables. While the size of the factor zoo gets a lot of attention (Cochrane, 2011), you cannot solve this omitted-variables problem by controlling for observables (Heckman, 1979).

I show how a researcher can use a special kind of error, called a "cognitive error", to identify whether investors are pricing assets based on a particular variable, X. Cognitive errors are instruments for identifying how mostly rational investors price assets.

Identification Framework

I start in Section 3 by writing down a framework for thinking about researchers' omitted-variables problem. A researcher wants to know whether investors are using a particular predictor, X, to price assets. He tests this hypothesis by regressing the cross-section of returns on the realized levels of X in some empirical setting:

$$r_n = \hat{a} + \hat{b} \cdot x_n + \hat{\varepsilon}_n \qquad \text{assets } n = 1, \ldots, N \tag{1}$$

All else equal, a higher current price implies a lower future return. So if investors prefer low X assets, a researcher should estimate $\hat{b} > 0$ in a well-identified setting.

Unfortunately, even if investors are not using X to price assets, the variable might still predict the cross-section of returns in a poorly identified empirical setting. This will occur whenever X happens to be correlated with some other variable, X', that investors do care about. If the researcher regresses the cross-section of returns on X in such a setting, he may spuriously estimate $\hat{b} > 0$.

The researcher wants to identify X as part of investors' true model because he wants to ensure that his results are robust to future discoveries. That is why he cares about identification. If he estimates $\hat{b} > 0$ in a well-identified setting, then he can be sure that investors will prefer low-X assets in every other setting too since investors always use the same model to price assets. By contrast, if the researcher estimates $\hat{b} > 0$ in a poorly identified setting, then he is liable to have his finding explained away by some future researcher in a new empirical setting where X and X' are no longer correlated.

Any confounding explanation for the return predictability associated with X must take the form of a correlated variable. So, to identify whether investors are using X to price assets, the researcher must find an empirical setting where X is uncorrelated with every other variable that investors know about. If he were to estimate $\hat{b} > 0$ in such a setting, then the researcher could be sure that investors were using X to price assets. Investors would not know about any other omitted variables in that setting.

One way to do this would be to construct test assets that only differed in X. This approach is analogous to flipping a coin to see which patients get a new drug, $X = 1$, and which get a sugar pill, $X = 0$. Another way to do this would be to look for a pre-existing subset of market data in which there are good reasons to believe that assets only differ in

X. This approach is like finding an instrumental variable. Due to the Vietnam draft lottery (Angrist, 1990), there are good reasons to believe that date of birth was the only difference between particular subset of Army veterans, $X = 1$, and an otherwise similar group of young men born one day earlier who did not enlist, $X = 0$. Both approaches are ways to ensure that $\mathbb{C}\text{ov}[X, X'] = 0$ for all other predictors X'.

Whichever approach the researcher takes, he needs to make sure that X is uncorrelated with all other predictors that investors know about not just the ones he is currently aware of. The researcher is worried about his finding being explained away by some future researcher using an as-yet-undiscovered predictor. So he must do more than consider the predictors he already knows about.

Cognitive Errors

But how can a researcher say anything useful about predictors which he himself has yet to discover? The key insight in this paper is that, when otherwise intelligent investors fail to correct an error, the researcher can learn something about what these investors did not know. The investors must not have known anything which would have allowed them to spot their mistake. If they had, they would have stopped making it. Crucially, this logic reveals what investors do not know regardless of what researchers do know.

In Section 4, I define a special kind of error, called a "cognitive error", which investors can only correct if they are aware of at least one such omitted variable. Let $\tilde{X}$ denote the version of X that investors perceive when making a cognitive error. If investors fail to correct a cognitive error about X in some empirical setting, a researcher can conclude that these investors must not know about any omitted variables. And I show that, if X and $\tilde{X}$ both predict returns in the same way in that setting, $\hat{b} > 0$, then the researcher can be absolutely sure that investors are using X to price assets. Cognitive errors are instruments for identifying how mostly rational investors price assets.

Section 5 gives examples showing how researchers can use cognitive errors as instruments for identifying how mostly rational investors price assets. I show how a quirk in how S&P/Barra classified stocks as "value" or "growth" results in a cognitive error that can be used to identify whether the value premium exists because investors price value and growth stocks differently (Boyer, 2011). I also show how peer effects (Bailey, Cao, Kuchler, and Stroebel, 2018) and investor inattention (Birru, 2015) can produce cognitive errors that are useful for identification purposes.

Efficient Markets

In efficient markets, "if there's a signal not now incorporated in prices that future values will be high, traders buy on that signal... bidding up the price until it fully reflects available information. (Cochrane, 2013)" The Efficient Market Hypothesis (Fama, 1970) claims that investors are able to correct every error that they initially make before a researcher sees the effects in his data.

The conventional wisdom is that it would be easiest for a researcher to identify investors' true model if this claim were true. In Section 6, I build on the above logic to show that the conventional wisdom has it exactly backwards: the assumption of market efficiency makes it impossible for a researcher to identify investors' model.

For markets to be efficient, investors must be able to correct every kind of error, cognitive errors included. But, to correct a cognitive error about X, investors must know about at least one correlated predictor. And, because correlated predictors are potential confounds, when a researcher assumes that markets are efficient, he is assuming that investors must know about a potential confound for any $\hat{b} > 0$ he estimates.

Suppose researchers are not aware of a potential confound. No matter. When they assume markets are efficient, they assume that investors must be. The assumption of market efficiency guarantees that investors always know about a problematic omitted variable. The only remaining question is how long before a future researcher discovers it.

1.1 Related Literature

The factor zoo literature (e.g., Harvey, Liu, and Zhu, 2016; Bryzgalova, 2017; Kozak, Nagel, and Santosh, 2018; Feng, Giglio, and Xiu, 2020; Freyberger, Neuhierl, and Weber, 2020) analyzes the implications of having lots of observed confounds. I point out that, when trying to identify investors' model, a researcher must also consider the omitted variables coming down the research pipeline.

Lucas (1976) shows how the results of a policy experiment may be distorted if investors strategically change their behavior in anticipation of the shock. Chemla and Hennessy (2019) and Hennessy and Strebulaev (2019) study different ways this can bias parameter estimates. Martin and Nagel (2019) describes how tests of market efficiency must change when investors face the same high-dimensional problem as researchers. These papers are all about estimation. The focus in this paper is identification.

Chemla and Hennessy (2021a,b) study how equilibrium interferes with identification. These papers take the underlying economic model as given. Then, the authors study the identification problem that emerges when a researcher interprets his empirical results when conditioning on this model being at equilibrium. I am studying a different identification problem. The problem in this paper stems, not from a researcher conditioning on a specific economic model being in equilibrium, but from the researcher no knowing which economic model to use. Asset-pricing researchers in particular have substantial leeway along this dimension (Chen, Dou, and Kogan, 2021).

There are many different frameworks for thinking about identification: potential outcomes (Rubin, 1974; Imbens and Angrist, 1994), structural equations (Haavelmo, 1943; Heckman and Vytlacil, 2005), causal diagrams (Wright, 1934; Pearl, 2009), and distributional robustness (Shimodaira, 2000; Peters, Bühlmann, and Meinshausen, 2016). This paper does not offer yet another one. Instead, it shows how a researcher can exploit a special kind of investor error to satisfy the key assumptions in all of them.

Papers like Altonji, Elder, and Taber (2005) and Oster (2019) show how, under certain conditions, a researcher can check for omitted variables by examining parameter stability given different sets of control variables. When an asset-pricing researcher assumes markets are efficient, these conditions will never be met.

Finally, several recent papers have explored optimal experimental design. Kasy (2016) and Banerjee, Chassang, Montero, and Snowberg (2020) both examine how a researcher should set up his experiment to get results with robust out-of-sample performance. There is also an active literature looking at the costs and benefits of randomized controlled trials (Deaton, 2010; Imbens, 2010; Olken, 2015). In the context of asset-pricing theory, an random variation in investor beliefs must represent an error. I point out that such errors are the cost of learning investors' model.

2 Canary Traps

Why does it have to be an error? Investors are often more intelligent and better informed than the researchers who study them. These are core assumptions that every asset-pricing researcher holds dear. So it might at first seem a bit odd to rely on investor errors for identification.

But identifying the effect of a strategic choice requires observing some bad choices.[1] To identify which model investors are using, a researcher needs these investors to be making a few errors precisely because he believes them to be intelligent and well informed. If investors to be uninformed simpletons, the researcher would not need any errors to identify how they were pricing assets. The pattern would be patently obvious given enough observational data.

To see why, imagine you are running the CIA, and you are concerned that one of your spies is leaking information. Each of your spies is intelligent; that is why you hired them. Each of your spies also knows things you do not; they are out in the field while you are sitting at a desk in Langley, VA.

In spite of these disadvantages, one way you could identify the mole would be to set a "canary trap" (Clancy, 1988). The next time you pass on a top-secret document, give a slightly different version to each spy. If you plant unique incorrect details in each version, then when your enemy acts on the unique brand of bad information he has received, you will immediately be able to identify which spy gave it to him.

This metaphor highlights two key points. First, the canary trap is necessary precisely because you want to identify an intelligent better-informed agent. If the mole in your network always followed rote rules of thumb and you knew everything he did and more, then setting a canary trap would be overkill. Why go through all the trouble? In the same way that you do not need canary traps to catch ham-fisted spies, a researcher does not need cognitive errors to identify how dumb investors are pricing assets.

Second, error plays a critical role. Even if this mole knows things you do not, when he fails to notice the incorrect details in his document, you can be sure that he does not know any corroborating details which would have allowed him to recognize the mistake. As a result, you can be confident that the suspected traitor's behavior does not merely look incriminating to you because you lack context. Likewise, when investors fail to correct a cognitive error about X, a researcher can be sure that he does not merely think that X is priced because he lacks context. He can be sure that investors are not aware of some as-yet-undiscovered omitted variable which could also explain the result.

[1] Flipping a coin to see who gets a new drug ensures that half the patients in a clinical trial receive suboptimal care. The efficacy of the new drug merely determines which half. When studying the effect of army service on future income, Angrist (1990) compared pairs of otherwise identical people, only one of whom enlisted. It cannot be optimal for two truly identical people to make different enlistment decisions.

3 Identification Framework

This section introduces a framework for thinking about an asset-pricing researcher's omitted-variables problem. There are two stages. First, a group of investors prices $N \gg 1$ risky assets using some asset-pricing model. Then, a researcher tries to identify whether the model these investors were using included a particular predictor, X. The omitted-variables problem stems from the fact that investors know about predictors which the researcher has yet to discover.

3.1 Empirical Setting

Investors pay p_n dollars to buy a share of the nth risky asset, which entitles them to a payout of v_n dollars at the end of the first stage. The realized return is $r_n = (v_n - p_n)/p_n$.

There are $K \gg 1$ variables that might forecast future payouts. Let $\mathcal{K}$ denote the set of all such predictors with $K = |\mathcal{K}|$. Let X_k denote one of these predictors, such as book-to-market ratio, exposure to labor-income risk, sales growth, returns over the past 6 months, etc. I use $x_{n,k}$ to denote the realized level of this predictor for the nth risky asset. I use Z to denote a nuisance variable which never affects future payouts.

Investors are aware of all K candidate predictors, $\mathcal{K}^I = \mathcal{K}$. At the start of the first stage, they calculate the level of every predictor for each asset, $x_{n,k}$. Investors are unaware of the nuisance variable, Z, and do not calculate its realized values.

Investors are aware of how each asset's future payout will be determined, $v \leftarrow \mathbb{V}(X_1 = x_1, \ldots, X_K = x_K, Y = y)$, as a function of the predictor levels that they observe and an unobserved random shock, y. They also have a particular asset-pricing model in mind, $p \leftarrow \mathbb{P}(X_1 = x_1, \ldots, X_K = x_K)$. This model specifies how an asset will be priced at the start of the first stage as a function of its predictor levels. This payout rule and asset-pricing model together with distributional assumptions concerning predictors, the nuisance variable, and the unobserved payout shock define a structural model.

Definition 1 (Entailed Distribution)**.** *Let π denote the probability distribution function entailed by this structural model in a particular empirical setting. This PDF is defined over payouts, prices, returns, all predictors $\mathcal{K}$, and the nuisance variable.*

Let π^I denote the marginal distribution entailed by the same structural model given investors' information set after assets are priced but before payouts are realized. This PDF is defined over prices and predictors.

In the second stage, the researcher observes the price p_n, payout v_n, and return r_n of each risky asset. The researcher is only aware of a subset of the predictors that investors know about, $\mathcal{K}^R \subset \mathcal{K}^I$. He sees the realized level of each of these predictors for every risky asset. The researcher is aware of the nuisance variable and observes z_n. He knows that there is some noise in each asset's payout. But he does not see the realized value, y_n.

In some ways, the researcher is less informed than investors. For example, investors know about more predictors than the researcher does. Investors also know which model they are using; whereas, the researcher is trying to learn this model from the data.

But the researcher also knows some things that investors do not at time when they are pricing assets. For example, since the researcher is looking at past data, the researcher can see each asset's realized payout. The researcher is also aware of the nuisance variable, Z, which sits below investors' waterline of cognition.

Definition 2 (Empirical Density)**.** *Let $\hat{\pi}$ denote the empirical density of payouts, prices, returns, all predictors $\mathcal{K}$, and the nuisance variable in a particular empirical setting.*

Let $\hat{\pi}^R$ denote the marginal density in that setting given the researcher's information set in the second stage. This density is defined over payouts, prices, returns, the subset of predictors the researcher knows about $\mathcal{K}^R$, and the nuisance variable.

Let $\hat{\pi}^I$ denote the marginal density in the same setting given investors' information set after assets are priced but before payouts are realized. This density is defined over prices and all predictors investors know about $\mathcal{K}^I = \mathcal{K}$.

As Manski (2009, p2) explains: "Studies of identification seek to characterize the conclusions that could be drawn if one could use the sampling process to obtain an unlimited number of observations... Statistical inference problems may be severe in small samples but diminish in importance as the sampling process generates more observations. Identification problems cannot be solved by gathering more of the same kind of data." To emphasize this point, I assume that N is large enough that sampling error can be ignored. There will be no meaningful difference between the results of calculations such as $\hat{\mathbb{E}}[\boldsymbol{x}_k] = \frac{1}{N} \cdot \sum_{n=1}^{N} x_{n,k}$ and $\mathbb{E}[X_k] = \int X_k \cdot \mathrm{d}\pi$.

Assumption A (N Is Arbitrarily Large)**.** *The number of risky assets, N, is large enough that you can ignore the difference that results from using the empirical density, $\hat{\pi}$, rather than the entailed distribution, π, to calculate any mean, variance, or covariance.*

The researcher wants to know whether the investors in his empirical setting were using some predictor to price assets. I will use $k = 1$ to denote this predictor of interest. The researcher tests his hypothesis by regressing the cross-section of returns on X_1:

$$r_n = \hat{a} + \hat{b} \cdot x_{n,1} + \hat{\varepsilon}_n \qquad \text{assets } n = 1, \ldots, N \tag{2}$$

All else equal, a higher current price implies a lower future return. So if investors prefer assets with lower X_1 values, then the researcher should estimate $\hat{b} > 0$ in an empirical setting where assets only systematically differ in their values of X_1.

3.2 Market Environment

A market environment corresponds to a specific country, time period, asset class, etc. Let $\mathcal{E}$ denote the set of all possible market environments. The defining feature of market environment e is the cross-sectional distribution of predictors in that environment:

$$\mathbb{C}\mathrm{ov}[X_1, X_2, \ldots, X_K] = \begin{pmatrix} 1 & \sigma_{1,2}(e) & \cdots & \sigma_{1,K}(e) \\ \sigma_{2,1}(e) & 1 & \cdots & \sigma_{2,K}(e) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K,1}(e) & \sigma_{K,2}(e) & \cdots & 1 \end{pmatrix} \tag{3}$$

For example, value stocks might tend to have lots of exposure to labor-income risk in US data but not in Japanese data.

Without loss of generality, all predictors are normalized to have $\mathbb{E}[X_k] = 0$ and $\mathbb{V}\mathrm{ar}[X_k] = 1$ in every market environment. The nuisance variable is such that $\mathbb{E}[Z] = 0$, $\mathbb{V}\mathrm{ar}[Z] = 1$, and $\mathbb{C}\mathrm{ov}[X_k, Z] = 0$ for all $k \in \mathcal{K}$ in every market environment.

The omitted-variables problem at the heart of this paper stems from the existence of predictors that are known to investors but not to researchers, $\mathcal{K}^R \subset \mathcal{K}^I$. These predictors are the eponymous omitted variables. You cannot solve an omitted-variables problem by controlling for observables (Heckman, 1979). So I assume that the researcher has already controlled for all observed confounds in his empirical setting.

Assumption B (No Observed Confounds)**.** *X_1 is uncorrelated with all other predictors that the researcher currently knows about,* $\mathbb{C}\mathrm{ov}[X_1, X_k] = 0$ *for all* $k \in \mathcal{K}^R \setminus 1$.

Since $N \gg 1$ is sufficiently large, it is always possible to orthogonalize the observed values of $\boldsymbol{x}_1 = \{x_{n,1}\}_{n=1}^N$ relative to every other predictor the researcher knows about. Assumption B basically says that the researcher has already taken this step.

3.3 Asset-Pricing Model

Each risky asset's future payout is determined as follows:

$$v \leftarrow \mathbb{V}(X_1 = x_1, \ldots, X_K = x_K, Y = y) = \mu_V - \sum_{k \in \mathcal{K}} \phi_k \cdot x_k + y \tag{4}$$

$\mu_V = \mathbb{E}[V]$ is the average risky-asset payout.[2] Each slope coefficient, $\phi_k \geq 0$, captures the effect of an increase in the kth predictor on an asset's future payout. I use the convention that investors dislike higher predictor values, hence the negative sign in Equation (4). For instance, this would be the case if the kth predictor represented a risk-factor exposure.

$Y \sim \text{Normal}[0, \sigma_Y^2]$ represents payout noise that is uncorrelated with every predictor: $\mathbb{C}\text{ov}[Z, Y] = 0$ and $\mathbb{C}\text{ov}[X_k, Y] = 0$ for all $k \in \mathcal{K}$. I will be writing σ_Y^2 rather than $\sigma_Y(e)^2$ for clarity. However, the variance of Y must differ across market environments to ensure a constant payout variance, $\sigma_V^2 = 1$.

To ensure researchers face a meaningful identification problem, some of the predictors must be irrelevant for forecasting future payouts. If this were not the case, then it would not be possible for a researcher to choose the wrong predictors.

Assumption C (Sparse Payout Rule)**.** $\phi_k = 0$ *for all but* $A \ll K$ *active predictors. Let* $\mathcal{A} = \{\, k \in \mathcal{K} : \phi_k > 0 \,\} \subset \mathcal{K}$ *denote the active subset of predictors.*

In addition, something about the payout rule must remain constant across market environments. Otherwise a researcher would not be able to apply insights from one market environment when making a prediction in another. Extrapolation is the whole point of identification (Manski, 1995). So I assume that the same payout coefficients are at work in all market environments.

[2] To avoid having to worry about risky-asset prices that are negative or zero, which would make returns ill-defined, I assume that the average risky-asset payout satisfies $\mu_V > \sqrt{2 \cdot \log N}$. This choice comes from the scaling relationship $\mathbb{E}\{\max_{n=1,\ldots,N} |\xi_n|\} \propto \sqrt{2 \cdot \log N}$ when $\xi_n \overset{\text{IID}}{\sim} \text{Normal}[0, 1]$.

Assumption D (Fixed Payout Rule)**.** *The future payouts of the N risky assets in every market environment are governed by the same coefficients,* $\{\mu_V, \phi_1, \ldots, \phi_K\}$.

The investors in every market environment use the same linear rule to price the risky assets:

$$\mathbb{P}(X_1, \ldots, X_K) = \lambda_0 - \sum_{k \in \mathcal{K}} \lambda_k \cdot X_k \tag{5}$$

Most asset-pricing models generate linear pricing rules of this sort. See Appendix B for one example. The negative sign in Equation (5) stems from the convention that investors dislike higher predictor values. I associate investors' asset-pricing model with the correct choice of pricing-rule coefficients implied by this model.

Definition 3 (Asset-Pricing Model)**.** *Each asset-pricing model m is defined by a collection of pricing-rule coefficients,* $\{\lambda_0, \lambda_1, \ldots, \lambda_K\}$*. Let $\mathcal{M}$ denote all such models.*

3.4 Model Implementation

Investors may fail to implement their model correctly for two different reasons. First, investors could apply the wrong coefficients. For example, suppose that investors under-react to negative earnings surprises, X_k. When these investors see a high realization of X_k, they do not reduce an asset's price enough. In this scenario, investors would be using $\tilde{\lambda}_k < \lambda_k$. Let $\Delta[\lambda_k] = \tilde{\lambda}_k - \lambda_k$ denote the difference between the pricing-rule coefficients that investors are using and the correct choice of coefficient implied by their asset-pricing model.

Second, investors could plug in the wrong inputs. The nth asset realizes $X_k = x_{n,k}$ for the kth predictor; however, suppose that investors calculate a value of $\tilde{x}_{n,k}$ for that asset. If $\tilde{x}_{n,k} = x_{n,k}$, then investors use the correct level for that predictor. However, if investors misread their data or make a coding error, then they may calculate the wrong level of some predictor $\tilde{x}_{n,k} \neq x_{n,k}$. Let $\Delta[x_{n,k}] = \tilde{x}_{n,k} - x_{n,k}$ denote the difference between the level of the kth predictor for the nth asset in investors' eyes and its actual level.

For example, financial news sites typically report the S&P 500's price growth, $\mathrm{d}\log p_{S\&P500}$, not its realized return, $r_{S\&P500}$. And Hartzmark and Solomon (2020) document that investors often fail to recognize the difference. Investors who made this error would price the S&P 500 incorrectly because they applied the pricing-rule coefficient for past returns to past price growth, $\tilde{x} = \mathrm{d}\log p_{S\&P500} \neq x = r_{S\&P500}$.

Or, consider Kansas City Value Line (KCVL) futures contracts. Prior to 1988, their payout was a function of the average price growth across stocks. Unfortunately, many investors mistakenly priced KCVL futures using the wrong kind of average (Ritter, 1996; Thomas, 2002; Mehrling, 2011). They calculated $\tilde{x}$ = arithmetic average price growth instead of x = geometric average price growth.

Definition 4 (Model Implementation)**.** *An implementation, i, is defined by an error operator, Δ^i, which describes how the pricing-rule coefficients used by investors and the realized predictor levels that investors calculate differ from the correct values. I use $i = \star$ to denote the rational benchmark:*

$$\Delta^{\star}[\lambda_k] = \Delta^{\star}[x_{n,k}] = 0 \tag{6}$$

Any other implementation requires investors to make at least one error.

Investors are worse off whenever they deviate from this rational benchmark. Using the wrong coefficients and/or applying the right coefficients to the wrong predictor levels will distort their portfolio holdings. And there is no way to improve investors' welfare by adjusting their portfolio holdings away from the rational-expectations equilibrium. This is why investors would prefer to implement their asset-pricing model correctly.

3.5 Omitted-Variables Problem

The researcher sees the data produced by investors who priced assets in a particular empirical setting. The researcher does not know which model the investors used. Unfortunately, the researcher also does not know about all of the predictors in investors' information set. Thus, the researcher can only assess whether the empirical setting he observes could, in principle, be consistent with different kinds of models.

Definition 5 (Consistent Models)**.** *The researcher's empirical setting is consistent with model m under implementation i if the empirical density $\hat{\pi}^R$ is covariance equivalent to the PDF defined over the same variables that is entailed by implementing model m in some empirical environment, $e \in \mathcal{E}$. Let $\mathcal{M}^R(i) \subseteq \mathcal{M}$ denote the set of all models consistent with the researcher's data under implementation $i \in \mathcal{I}$.*

$\mathcal{M}_1^R(i) = \{m \in \mathcal{M}^R(i) : \lambda_1 > 0\}$ *is the subset of all consistent models under implementation i that include the predictor that the researcher is interested in, X_1.*

The researcher's empirical setting identifies X_1 as part of investors' true model if X_1 is either active in every consistent model or inactive in every consistent model.

Definition 6 (Identified Setting)**.** *The researcher's empirical setting identifies whether X_1 is part of investors' true model under implementation i if every consistent model contains predictor X_1 or no consistent model contains X_1:*

$$\mathcal{M}_1^R(i) = \begin{cases} \mathcal{M}^R(i) & \text{if } \lambda_1 > 0 \\ \emptyset & \text{if } \lambda_1 = 0 \end{cases} \tag{7}$$

Lemma 1 (Identified Setting $\Leftrightarrow$ No Correlated Omitted Variables)**.** *A setting identifies whether X_1 is part of investors' model under implementation i if and only if investors do not know about any correlated omitted variables:* $\mathbb{C}\mathrm{ov}[X_1, X_k] = 0$ *for all* $k \in \mathcal{K}^I \setminus \mathcal{K}^R$.

If there exists at least one identified setting under implementation i, then it is possible (at least in principle) for a researcher to identify whether investors are using X_1 to price assets. But it may be possible that no identified setting exists depending on what the researcher assumes about how investors implement their chosen model.

Below I give a formal definition of the omitted-variables problem that the researcher is trying to solve. Here is the idea. Suppose that investors use some model $m \in \mathcal{M}$ to price risky assets in some market environment $e \in \mathcal{E}$ under implementation i. This will entail a probability distribution. A researcher observes the corresponding empirical density for a subset of variables, $\hat{\pi}^R$. Based on this information and the correct assumption about how investors implemented their model, the researcher must decide whether the setting identifies X_1 is part of investors' model.

Definition 7 (Omitted-Variables Problem)**.**

> **(Setting)** *An empirical density, $\hat{\pi}^R$, that was generated by investors using a particular model $m \in \mathcal{M}$ to price the risky assets in a specific market environment $e \in \mathcal{E}$ under implementation i.*
>
> **(Query)** *Does the setting identify whether X_1 is part of investors' model?*

An asset-pricing researcher can answer with "yes", "no", or "I do not know". His answer is correct if he says "yes" when the empirical setting is identified or "no" when it is not. His answer is incorrect if he says "yes" when the setting is unidentified or "no" when it is identified. An identification strategy is a rule that tells the researcher which answer to give for a particular setting.

Definition 8 (Valid Identification Strategy)**.** *A valid identification strategy never gives an incorrect answer in any setting and correctly answers "yes" in at least one setting.*

There are many different kinds of identification strategies. None of them works in all settings. However, each of them gives conditions such that, if a particular empirical setting satisfies those conditions, the researcher can conclude that the setting is identified. For example, instrumental variables is a valid identification strategy. Not every market setting contains an instrument, but some do. And the potential-outcomes framework lays out criteria that need to be satisfied by any such instrument. If those criteria are satisfied, then the researcher will never falsely think that his empirical setting is identified.

Small stocks tend to be more illiquid in this world, $\mathbb{C}\mathrm{ov}[\mathit{LIQ}, \mathit{SIZE}] > 0$. Suppose that investors are pricing assets based on liquidity (Amihud, 2002) but not size (Banz, 1981), and consider an asset-pricing researcher living in 1994. This researcher exists at a time after the size effect has been widely publicized but before liquidity has gained traction in the academic research. So, even though liquidity and size are highly correlated, the researcher has never thought to use liquidity as a predictor.

This researcher regresses the cross-section of returns on *SIZE*. Since small stocks tend to be more illiquid, the researcher finds evidence of a size effect, $\hat{b} > 0$. But the researcher also knows that there will be future discoveries coming down the pipeline. So he recognizes that his results might be due to an omitted variable. He does not yet know about *LIQ* as a predictor in 1994, but he does recognize that future researchers might discover a variable, such as liquidity, which seems to account for much of the size effect.

To solve this researcher's omitted-variables problem, you must come up with a valid identification strategy that the researcher could use in 1994 to determine whether or not investors were pricing assets based on *SIZE*. Since we are assuming that the researcher has already orthogonalized *SIZE* with respect to every observable predictor in each empirical setting, this amounts to giving conditions under which the researcher can be sure that *SIZE* is uncorrelated with every other as-yet-undiscovered predictor that investors know about. If the researcher encounters such a setting, he will be able to identify that *SIZE* is not part of investors model.

4 Main Result

This section describes how to use a special kind of investor error, called a "cognitive error", to solve this omitted-variables problem. The trick is to define a cognitive error about X_1 in such a way that otherwise intelligent investors will only fail to recognize their error in empirical settings where they do not know about any correlated omitted variables. That way, when a researcher encounters an empirical setting where investors fail to correct a cognitive error about X_1, he can be absolutely sure that these investors are not aware of any potential confounds—i.e., that the setting is identified.

4.1 Key Assumption

The logic outlined above hinges on one key assumption: while investors sometimes make mistakes, they are not completely foolish. Investors are aware that they could be implementing their model incorrectly. They are aware that they may be using some $\Delta^i \neq \Delta^\star$. However, they do not initially know the nature of their mistake.

After assets are priced but before payouts are realized, investors get a chance to correct their initial error. Investors know how prices and predictors should be distributed under the rational benchmark, $\pi^I(\star)$. Their economic intuition tells them that they should observe this entailed distribution. Investors can also inspect the empirical density of prices and predictors for the N risky assets in their setting, $\hat{\pi}^I(i)$.

Investors look for evidence that they are using some implementation $\Delta^i \neq \Delta^\star$ by comparing this entailed distribution, $\pi^I(\star)$, to the observed empirical density, $\hat{\pi}^I(i)$. Since N is arbitrarily large, there will be no mismatch between these two distributions under the rational benchmark, $i = \star$. In essence, investors ask themselves: "If I take my model seriously, then I should see such and such. When I look at the way I just priced assets, do I see that this prediction actually holds up?"

Assumption ♠ (Investors Are Otherwise Intelligent)**.** *Investors recognize that $\Delta^i \neq \Delta^\star$ in every empirical setting where the entailed distribution, $\pi^I(\star)$, and the empirical density, $\hat{\pi}^I(i)$, are not covariance equivalent given investors' information set.*

Whenever the investors in a particular empirical setting recognize that they were using $\Delta^i \neq \Delta^\star$, they implement their model using the $i = \star$, and the researcher sees market data that were produced under the rational benchmark in the second stage.

Like in Tirole (2009), if the entailed distribution, $\pi^I(\star)$, and the empirical density, $\hat{\pi}^I(i)$, are not covariance equivalent given investors' information set, it is an "eye opener". Investors immediately recognize their error and adjust prices to reflect the rational benchmark. A researcher studying an empirical setting in which investors have recognized and corrected an initial implementation error will never see any of its effects.

Otherwise intelligent investors may not always be able to spot an implementation error, though. This will happen whenever investors' information set is not fine enough to distinguish between $\pi^I(\star)$ and $\hat{\pi}^I(i)$. The key insight in this paper is that, when this happens, it reveals something about what investors did not know. They must not have known about anything which would have allowed them to tell that $\pi^I(\star)$ and $\hat{\pi}^I(i)$ were not covariance equivalent. Importantly, this constraint on what investors must not have known is independent of what researchers currently do know.

4.2 Cognitive Errors

I operationalize this insight by defining a special kind of error about X_1, called a "cognitive error", that investors only fail to recognize when they are not aware of any correlated omitted variables. In other words, a cognitive error is defined in such a way that investors could use any omitted variable that a researcher would classify as a confounding explanation for error correction purposes.

Definition 9 (Cognitive Error)**.** *When investors make a cognitive error about X_1, they implement their model using an error operator defined as $\Delta[\lambda_k] = 0$ for all $k \geq 0$,*

$$\Delta[x_{n,1}] = (\sqrt{1-\theta} - 1) \cdot x_{n,1} + \sqrt{\theta} \cdot z_n \qquad \textit{for some } \theta \in [0, 1) \tag{8}$$

and $\Delta[x_{n,k}] = 0$ for all $k > 1$. When investors recognize a cognitive error, they set $\theta = 0$.

When investors make a cognitive error about X_1, it is as if they are unknowingly anchoring on the nuisance variable Z à la Tversky and Kahneman (1974). Investors are not directly aware of this variable, but it still seeps into their decision-making process.

Larger values of θ correspond to more severe anchoring. In the limit as $\theta \to 1$, investors incorrectly calculate $\tilde{x}_{n,1} = z_n$ even though X_1 and Z are entirely unrelated, $\mathbb{C}\text{ov}[X_1, Z] = 0$. By contrast, if $\theta = 0$, investors are not making an error at all or they have recognized and corrected their mistake.

For example, when asked, home buyers claim that their beliefs about future house-price growth are not affected by their out-of-state friends' past experiences in the housing market. Nevertheless, their distant friends' housing-market experiences do distort their beliefs (Bailey, Cao, Kuchler, and Stroebel, 2018). When their distant friends' home prices have risen more, they unwittingly anchor their beliefs about future house-price growth in their own city on this good reference point.

Suppose that investors know about some other predictor, X_k, that is correlated with X_1, in their empirical setting, $\sigma_{1,k}(e) \neq 0$. If investors are making a cognitive error about X_1, then this error will create a mismatch between the covariance entailed by the rational benchmark, $\sigma_{1,k}(e)$, and the covariance they observe in their data:

$$\widehat{\mathbb{C}\text{ov}}[\tilde{\boldsymbol{x}}_1, \boldsymbol{x}_k] = \sqrt{1-\theta} \cdot \sigma_{1,k}(e) < \sigma_{1,k}(e) \tag{9}$$

A cognitive error about X_1 will deflate its covariance with any other correlated predictor. Thus, otherwise intelligent investors will only ever fail to correct a cognitive error about X_1 in empirical settings where they do not know about any correlated predictors.

Cognitive errors do not affect predictor variances, $\mathbb{V}\text{ar}[X_k] = 1$ for all k. Cognitive errors also do not affect any covariances that do not involve X_1. However, these two additional assumptions are not essential to the results in this paper. The key thing is that investors must be able to spot the error in any empirical setting where they are aware of at least one correlated predictor.

Suppose that a researcher knows that investors have failed to correct a cognitive error about X_1 in his empirical setting. The following proposition details how he can exploit this knowledge for identification purposes.

Proposition 1 (Valid Identification Strategy)**.** *If investors have failed to correct a cognitive error about X_1 in an empirical setting, then the setting is identified.*

Moreover, the following protocol represents a valid identification strategy. Regress the cross-section of returns on the Z associated with the cognitive error:

$$r_n = \hat{c} + \hat{d} \cdot z_n + \hat{\varepsilon}_n \tag{10}$$

X_1 is part of investors' model if $\hat{d} > 0$. It is not part of investors model if $\hat{d} = 0$.

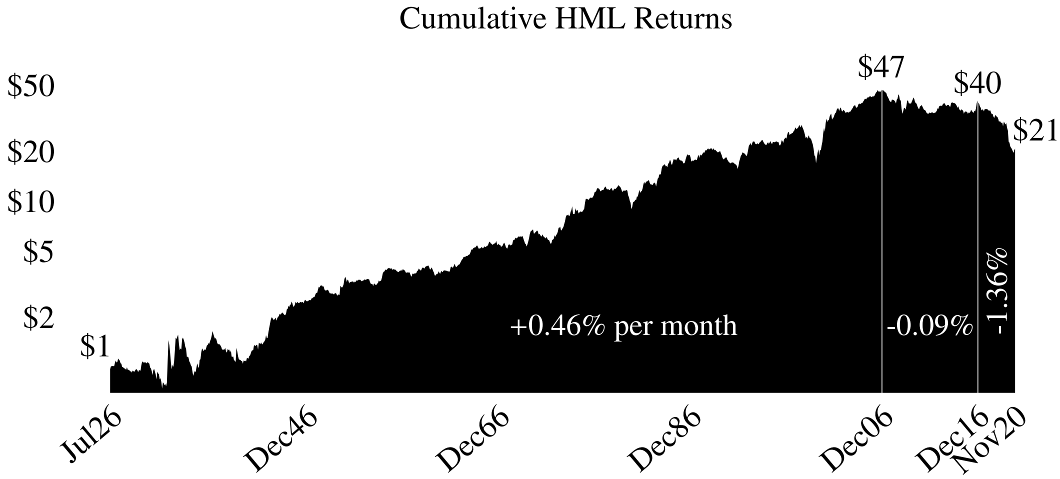

Figure 1. Cumulative returns to Fama and French (1993)'s high-minus-low (HML) book-to-market factor, F_{HML}, using monthly data from July 1926 to November 2020. y-axis is on a logarithmic scale. Percentages in white are average monthly returns to the HML factor during different time windows.

Notice that the logic behind this identification strategy converts a portion of the existing behavioral-finance literature into a cache of instruments. Researchers should not view the behavioral literature as an attack on rational models. Instead, this literature is a collection of tools for identifying how mostly rational investors price assets.

Cognitive errors offer a way to halt the explanation cycling—size matters (Banz, 1981); no, the size effect is dead (Berk, 1995); actually, size still matters when controlling for asset quality (Asness, Frazzini, Israel, Moskowitz, and Pedersen, 2018); etc—that has slowed progress in the field. Once a researcher uses a cognitive error to identify X_1, he no longer has to worry his result being explained away by an omitted variable that will be discovered by some future researcher.[3]

5 Some Applications

Do investors price value stocks differently than growth stocks because they care about the value-vs-growth distinction (Subsection 5.1)? Do home buyer beliefs about future house-price growth affect how these investors price homes (Subsection 5.2)? Are momentum profits the result of investors making pricing decisions based on realized

[3]This does not mean that the researcher's results would be infallible. In practice, the researcher would still have to worry about estimation error, mis-specification, etc. N is not arbitrarily large in the real world.

gains (Subsection 5.3)? This section shows how to answer these questions using cognitive errors taken from the existing behavioral-finance literature.

Along the way, I point out promising places to look for more cognitive errors. But for the same reasons that there is no mechanical recipe for finding instrumental variables (Angrist and Krueger, 2001), I can offer no such recipe for finding cognitive errors. Finding instruments involves creativity. So does finding cognitive errors.

5.1 Label Changes

Does the value premium exist because investors prefer growth stocks? This subsection shows how to answer this question using a cognitive error taken from Boyer (2011). The cognitive error is based on a quirk in how S&P/Barra constructed its value- and growth-stock indexes.

Identification Problem

A portfolio that is long high book-to-market (B/M) stocks and short low B/M stocks (Fama and French, 1993) has an average excess return of $\hat{b} = 0.44 \pm 0.8\%$ per month during the period from July 1926 to November 2020 as shown in Figure 1.[4] This finding is consistent with the hypothesis that investors are willing to pay more for growth stocks.

But growth stocks also tend to be younger less-profitable firms. Both of these variables represent potential confounds. Either of them could be the true explanation for the 44bps per month value premium in Figure 1.

You cannot solve this identification problem by simply controlling firm age, profitability, etc using more sophisticated econometric techniques. These methods allow you to control for the observed predictors in the existing academic literature. But you cannot solve an omitted-variables problem by better controlling for observables.

You also cannot solve this identification problem by writing down a fancy theory. The challenge is to identify what investors are actually doing not what they theoretically could be doing (Chinco, Hartzmark, and Sussman, 2020). If you tested an additional theoretical prediction and found that it was satisfied in the data, the result would be consistent with your new story. But it would not represent unique identifying evidence that could only be explained this way.

[4] The data come from Ken French's website. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html.

Cognitive Error

To identify whether investors care about the value-vs-growth distinction when pricing assets, a researcher must find an empirical setting in which stocks sometimes change from being labeled as value to being labeled as growth in investors' minds for reasons unrelated to any other variables that investors know about.

Boyer (2011) documents a cognitive error that can be used for this purpose. S&P/Barra calls any S&P 500 stock with a book-to-market (B/M) above a certain threshold a "value stock". They label every other S&P 500 stock as a "growth stock". When S&P/Barra's increases this threshold, a value stock with an unchanged B/M can get reclassified as a growth stock. The same thing in reverse can happen to growth stocks.

From 1992 to 2004, there were 54 instances where S&P/Barra reclassified an S&P 500 stock from value to growth or vice versa when its B/M ratio had actually moved in the opposite direction. Boyer (2011) shows that investors respond to these reclassifications in the same way as every other reclassification, $\text{Sign}[\hat{b}] = \text{Sign}[\hat{d}]$. It is as if investors' mental calculation of a stock's valueness or growthiness is anchored on S&P/Barra's uninformative label.

To construct an identified setting, collect two observations about each of the 54 firms that were reclassified due to a change in S&P/Barra's value-vs-growth threshold: one before the reclassification and one immediately after. The only difference between these two observations would be how investors perceive the firm's B/M in each instance.

Let Z denote an indicator variable such that:

$$Z = \sqrt{2} \times \begin{cases} +1 & \text{if post-reclassification from growth to value} \\ 0 & \text{if pre-reclassification} \\ -1 & \text{if post-reclassification from value to growth} \end{cases} \tag{11}$$

The factor of $\sqrt{2}$ ensures that $\mathbb{V}\text{ar}[Z] = 0.25 \cdot (\sqrt{2})^2 + 0.50 \cdot 0^2 + 0.25 \cdot (-\sqrt{2})^2 = 1$, and the parameter $\theta \in [0, 1)$ captures the salience of S&P/Barra's labeling to investors' mental calculations. Table 2 in Boyer (2011) documents that $\sqrt{\theta} \cdot Z \in \{-30, 0, +30\}$bps, which would imply a $\theta = 0.30^2/2 \approx 0.05$.

Results/Implications

Proposition 1 says to regress the cross-section of 108 return observations in this empirical setting on Z:

$$r_n = \hat{c} + \hat{d} \cdot z_n + \hat{\varepsilon}_n \qquad \text{observations } n = 1, \ldots, 108 \tag{12}$$

If the estimated slope coefficient from Equation (12) is non-zero, $\hat{d} > 0$, then this would identify that the value premium was due to investors' preference for growth stocks. If investors were aware of any confounding explanations, then they would have stopped anchoring their perceptions of valueness and growthiness on S&P/Barra's labels. Given such an identifying result, you would not have to worry about a future PhD student proposing an alternative explanation for your result.

Classification mistakes and arbitrary label changes are an important cache of cognitive errors for future researchers to mine. For example, there are numerous papers looking at how changes in mutual-fund style labels often do not reflect changes in their actual holdings (Shleifer, 1986; Barberis, Shleifer, and Wurgler, 2005; Cooper, Gulen, and Rau, 2005; Sensoy, 2009). Companies can also be mislabeled in uninformative ways (Cooper, Dimitrov, and Rau, 2001; Rashes, 2001; Hartzmark and Sussman, 2019). And analysts sometimes make important labeling errors (Chen, Cohen, and Gurun, 2019; Ben-David, Li, Rossi, and Song, 2020). In the past, researchers have viewed situations where this sort of uninformative labeling affects asset prices as demonstrations of market inefficiency. This paper shows they are also more than that; they represent opportunities to better understand how investors price assets.

5.2 Social Connections

Do home buyer beliefs about future house-price growth affect how these investors price homes today? This subsection shows how to answer this question using a cognitive error taken from Bailey, Cao, Kuchler, and Stroebel (2018) involving peer effects.

Identification Problem

Cities with higher current house prices also tend to have higher future house-price growth. For example, in November 2000, the price of a typical house in Detroit was \$150k; whereas, the typical price in Los Angeles was \$253k. Over the subsequent two

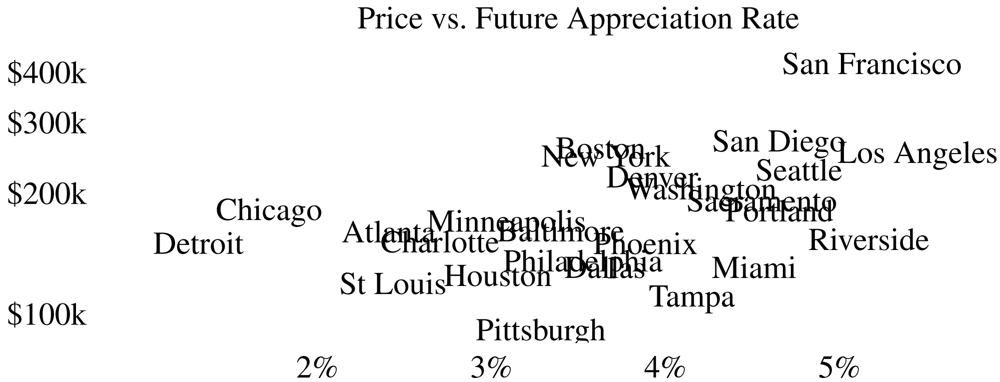

Figure 2. Relationship between the November 2000 house-price level in the largest 25 US cities (y-axis; log scale; dependent variable in Equation 13) and the house-price appreciation rate from December 2000 to November 2020 in these cities (x-axis; annualized; RHS variable in Equation 13).

decades, while house prices only grew at a rate of 1.32% per year in Detroit, they grew at a rate of 5.45% per year in Los Angeles.[5]

This pattern generalizes to other cities as illustrated by Figure 2. When I regress the log price level in each of the 25 largest US cities as of November 2000, $\log p_n$, on the annualized house-price growth rate over the next 20 years, x_n, I find that:

$$\log p_n \;=\; \hat{a} \;+\; \underset{[0.06]}{0.14} \cdot x_n \;+\; \hat{\varepsilon}_n \tag{13}$$

This regression says cities with 1% per year higher growth rates in the future tended to have current price levels that were $\hat{b} = 14 \pm 6\%$ higher on average.

This point estimate is consistent with the hypothesis that investors consider future house-price growth when deciding how much to pay for a home today. This makes intuitive sense. If the price of a house will rise in the future, whoever buys the house will pocket these gains. Knowing this, every potential buyer should be willing to pay a bit more for the house right now. Home buyers in standard models à la Poterba (1984) all think this way (e.g., see Case and Shiller, 2003; Himmelberg, Mayer, and Sinai, 2005).

However, home buyers in the real world do not have to think this way. The same sort of reasoning says that a homeowner who is currently underwater on his mortgage should default when the expected gains from future house-price growth are less than the present

[5]These data come from Zillow Research. See `https://www.zillow.com/research/data/`.

value of his future mortgage payments. And we know most homeowners do not exercise this default option correctly (Deng, Quigley, and Van Order, 2000), which suggests real-world investors are not guaranteed to think about the present value of future house-price changes like a textbook investor would.

Moreover, in the same way that risk-factor betas are not exogenous in the stock market (Campbell and Mei, 1993), home buyers' beliefs about future house-price growth are not exogenous either. We can all think of lots of reasons besides higher future house-price growth why you might want to pay more for a house in Los Angeles than in Detroit. In short, this does not represent an identified setting.

Cognitive Error

Solving this omitted-variables problem requires finding an empirical setting with variation in perceived future house-price growth rates that is unrelated to any other variable that home buyers might know about. Bailey, Cao, Kuchler, and Stroebel (2018) suggests a cognitive error based on social connections.

Suppose Alice and Bob are home buyers living in Phoenix. The only difference between these two otherwise identical people is that Alice has a bunch of Facebook friends in Los Angeles (high appreciation) while most of Bob's friends live in Detroit (low appreciation). Bailey, Cao, Kuchler, and Stroebel (2018, Table 10) shows that, due to the influence of her Facebook connections, Alice's beliefs about house-price growth in Phoenix will be higher than Bob's. This belief distortion represents a cognitive error. Differences in where Alice and Bob's Facebook friends live can have nothing to do with underlying housing-market conditions in Phoenix.

Let X_1 denote the correct belief about the future house-price growth rate in a particular city. Alice and Bob should both have beliefs of x_1 if they were not making any errors. Let $\tilde{x}_{n,1}$ denote the distorted beliefs about future house-price growth rates held by a particular home buyer, $n \in \{\text{Alice}, \text{Bob}, \text{Charles}, \text{Diane}, \ldots\}$.

Define Z to be the average house-price growth experienced by a home buyer's Facebook friends that live some other city. For Alice and Bob, this is the house-price growth experienced by their Facebook friends who live anywhere but Phoenix. Assume this anchor has been normalized to have mean zero and unit variance across buyers, $\mathbb{E}[Z] = 0$ and $\mathbb{V}\text{ar}[Z] = 1$. The parameter $\theta \in (0, 1)$ captures the strength of home buyers' Facebook connections. Higher θ means they check Facebook more often.

This cognitive error makes it possible to create an identified setting for testing whether home buyer beliefs about future house-price growth affect current prices. This empirical setting contains home buyers, like Alice and Bob, who live in the same place and only differ in the past house-price growth enjoyed by their Facebook friends living in other cities. Bailey, Cao, Kuchler, and Stroebel (2018) use this empirical setting to document that these social connections influence a home buyers' willingness to pay. However, by extension, the authors' results also imply something important about the model that home buyers use to price houses.

Results/Implications

Bailey, Cao, Kuchler, and Stroebel (2018, Table 9) reports the results of the following specification. The authors regress the log price paid by the nth home buyer on the house-price growth experienced by this home buyers' Facebook friends living in some other city, z_n. This is the exact specification in Proposition 1 written in terms of prices rather than realized returns. The estimated slope coefficient implies that, if Alice's Facebook friends in other cities had experienced 10% higher house-price growth, Alice would have paid $4.5 \pm 1.5\%$ more than Bob.

The authors interpret this as evidence that social connections affect housing-market outcomes. And it is. But it is also more than that. By showing that Facebook connections distort house prices via their effect on beliefs about future house-price growth rates, Bailey, Cao, Kuchler, and Stroebel (2018) also identify home buyers' beliefs about future house-price growth rates as an input to these investors' model. They confirm that home buyers' beliefs about future house-price growth directly affect home prices today.

The literature on peer effects and social connections represents a natural place to look for cognitive errors going forward (Shiller, 1984; Shiller and Pound, 1989; Kuchler and Stroebel, 2020; Hirshleifer, 2020). This literature is concerned about whether a result can be explained by homophily rather than social influence (McPherson, Smith-Lovin, and Cook, 2001; Jackson, 2010). To show that an effect is not just a case of birds of a feather flocking together, an asset-pricing researcher must show that there is no unobserved fundamental shock around which to organize a flock. In short, identifying that a change in beliefs was due to peer effects requires a cognitive error.

5.3 Investor Inattention

Do momentum returns stem from investors worrying about realized gains when computing prices? Birru (2015) uses an approach similar to the one formalized in this paper to answer this question. The cognitive error in Birru (2015) is based on investor inattention: investors who suffer from the disposition effect often fail to adjust their reference price for stock splits.

Identification Problem

The disposition effect is the observation that, all things being equal, investors tend to sell winner stocks too quickly and hold loser stocks too long (Shefrin and Statman, 1985). Suppose a stock realizes good news and its price rises. A disposition effect investor will immediately sell the stock, resulting in downward price pressure and predictably higher future returns as the stock's price gradually corrects upwards towards its fundamental value. Thus, due to the disposition effect, prices might consistently under-react. This would lead to a pattern of return predictability similar to momentum.

Researchers define winner and loser stocks by comparing a stock's current price, P, to the price at which the investor purchased the stock, Ref. If the current price is higher, an investor has experienced gains:

$$Gain = \frac{P - Ref}{P} \tag{14}$$

Since researchers cannot see the price at which each investor purchased a stock, they proxy for the reference price with the weighted-average stock price over the past five years (Grinblatt and Han, 2005).

Birru (2015, Table 10a) shows that the quintile of stocks with the highest gains (winners) had weekly returns that were 0.16 ± 0.08%pt higher than those of the loser quintile. This difference in weekly returns corresponds to an $52 \cdot 0.16\% \approx 8.1\%$ annual excess return on a portfolio that sells losers and uses the proceeds to buy winners.

This evidence is consistent with the hypothesis that the disposition effect generates return predictability. However, winner stocks are different from loser stocks in other ways besides their past price growth. To state the obvious, winner stocks realized good news while loser stocks realized bad news. Investors could be pricing assets based on this good news and not the disposition effect.

Cognitive Error

Identifying whether investors care about realized gains when pricing stocks requires finding an empirical setting with variation in realized gains that is unrelated to any other variables investors might know about. Birru (2015) points to a cognitive error that can be used for this purpose: inattention to stock splits.

In a classic 1-to-2 stock split, each share of a company's stock worth P dollars gets converted into two shares worth $P/2$ dollars each. Following a 1-to-κ split, a single share of stock worth P dollars gets converted into κ shares, each worth P/κ dollars for some $\kappa > 0$. When $\kappa > 1$, as is the case in any forward split, a past winner may no longer seem like a past winner to an investor who fails to adjust his reference price:

$$\widetilde{Gain} = \frac{P/\kappa - Ref}{P/\kappa} = Gain + (1 - \kappa) \cdot \frac{Ref}{P} \tag{15}$$

An identified setting would contain otherwise identical pairs of stocks, each of which experienced a recent split. Each pair of stocks would only differ in their gains as perceived by an inattentive investor due to the split factor, κ. Let $Z = (1 - \kappa) \cdot (Ref/P)$ denote the change in perceived gains due to the occurrence of a stock split. The parameter $\theta \in (0, 1]$ then reflects the proportion of investors who are inattentive to splits.

Results/Implications

Birru (2015) regresses the cross-section of returns on *Gain* using only data on stocks in the four weeks after a split. If investors are inattentive to splits, they will perceive $\widetilde{Gain}$ rather than *Gain*. And, since all the stocks recently split, $\widetilde{Gain}$ will generally be very different from $Gain_n$ in this sample, which means that there should be little relationship between the true gains, *Gain*, and subsequent returns, r_n. This is exactly what Birru (2015, Table 10b) reports.

This related result is consistent with the hypothesis that past gains are part of investors' model, but it does not quite represent an identified setting. Finding that A never occurs in the absence of B does not imply that B causes A. Nevertheless, the result does allow Birru (2015) to show that the disposition effect cannot be a main driver of momentum (Jegadeesh and Titman, 1993). While *Gain* does not predict future returns in the post-split sample, past returns does.

6 Efficient Markets

Investors do not make cognitive errors in most empirical settings. Finding a valid instrument takes creativity and intuition. The same is true for cognitive errors. Cognitive errors do not make identification easy. They merely offer one valid identification strategy where none previously existed.

Of course, cognitive errors are not the only way to solve researchers' omitted-variables problem. As suggested by Table 1 column (3) in row s, there may be other tricks that a researcher could use to identify X_1 as part of investors' model.

However, there is a sense in which cognitive errors are a necessary condition for the existence of any other valid identification strategy. As shown in row s', there may be empirical settings where the only way to identify whether investors are using X_1 to price assets is to use a cognitive error. But, in this section, I prove that it is not possible to identify X_1 in any empirical setting where investors are guaranteed to correct any cognitive error as indicated by the red dotted line in row s'' of Table 1.

The conventional wisdom is that it would be easiest to identify investors' true model if markets were efficient. But when you eliminate the empirical settings where investors might fail to correct a cognitive error in Table 1, what remains are the empirical settings where investors know about problematic omitted variables. I show that there is no way to solve researchers' omitted-variables problem in any of these empirical settings. Every row with a "Yes" in column (4) has a "No" in columns (2) and (3).

6.1 Necessary Condition

Cognitive errors are just one valid identification strategy that an asset-pricing researcher might use to solve his omitted-variables problem. There may be other valid identification strategies. That being said, in order for there to be any other valid identification strategy, otherwise intelligent investors must not be able to correct a cognitive error about X_1. Put differently, if the researcher assumes that investors must be implementing their model using the rational benchmark, then he cannot identify whether X_1 is part of investors' model.

Proposition 2 (Necessary Condition)**.** *If a researcher assumes that otherwise intelligent investors can correct any cognitive error about X_1 in his empirical setting, then no valid identification strategy exists in that setting.*

	Researcher can solve omitted-variables problem using a. . .		Investors can correct every error that they initially make
Empirical setting	cognitive error	nother strategy	
(1)	(2)	(3)	(4)
1	No	No	Yes
2	No	No	Yes
3	No	No	Yes
4	No	No	Yes
5	No	No	Yes
⋮	⋮	⋮	⋮
147	No	No	Yes
148	No	No	Yes
⋮	⋮	⋮	⋮
s	Yes	Yes	No
⋮	⋮	⋮	⋮
18,212	No	No	Yes
18,213	No	No	Yes
18,214	No	No	Yes
⋮	⋮	⋮	⋮
s'	Yes	No	No
⋮	⋮	⋮	⋮
72,999	No	No	Yes
⋮	⋮	⋮	⋮
~~s''~~	~~No~~	~~Yes~~	~~Yes~~
⋮	⋮	⋮	⋮

Table 1. Necessary Condition. Illustration of why cognitive errors are a necessary condition for solving the omitted-variables problem that asset-pricing researchers face. Column (1) indexes each empirical setting that a researcher might observe data on. Columns (2) and (3) indicate whether a researcher can use either a cognitive error or some other approach to identify whether investors are pricing assets based on X_1 in each setting. Column (4) indicates whether investors are able to correct every kind of error that they might initially make in each setting. The dotted red line through row s'' indicates that a researcher can never encounter that sort of empirical setting.

The logic behind this result is straightforward. Suppose that X_1 predicts the cross-section of returns in a researcher's empirical setting, $\hat{b} > 0$. Further suppose that X_k is a variable that investors know about but which researchers have yet to discover. Any time X_k happens to be correlated with X_1 in the researcher's empirical setting, $\mathbb{C}\mathrm{ov}[X_1, X_k] \neq 0$, this omitted variable would represent a confounding explanation for his initial result suggesting that investors were pricing assets based on X_1.

If X_1 predicts the cross-section of returns in the researcher's data set, then some of the models consistent with this empirical setting will contain this predictor, $\mathcal{M}_1^I(\star) \neq \emptyset$. However, if investors know about an omitted variable that is correlated with X_1, there must also be consistent models that exclude the predictor, $\mathcal{M}_1^I(\star) \subset \mathcal{M}^I(\star)$. Thus, there is no way to identify whether X_1 is part of investors' true model in this setting.

Notice that the logic behind this result hinges on the definition of an identified setting (Definition 6) not any particular identification strategy, such as the one in Proposition 1. Thus, it applies to any identification strategy that a researcher might use to identify X_1 as part of investors' model based on a cross-sectional regression.

Shannon (1948) tells us that correcting errors requires redundant information. A cognitive error about X_1 only distorts this variable's covariance with other predictors. So, to recognize a difference between $\hat{\pi}^I(i)$ and $\pi^I(\star)$, investors must know about at least one correlated omitted variable. This is the kind of redundancy that is required. And such an omitted variable would thwart any identification strategy.

6.2 Conventional Wisdom

In efficient markets, "if there's a signal not now incorporated in prices that future values will be high, traders buy on that signal... bidding up the price until it fully reflects available information. (Cochrane, 2013)" The Efficient Market Hypothesis (Fama, 1970) claims that investors are able to correct every error that they initially make before a researcher sees the effects in his data. If markets are efficient, then a researcher will only every observe data in empirical settings with a "Yes" in column (4) of Table 1.

Market efficiency is "an extreme null hypothesis. We do not expect it to be literally true. (Fama, 1970)" Yet "most economists view market efficiency as a useful benchmark. (Campbell, 2017)" The conventional wisdom is that it would be easiest for a researcher to identify investors' true model if this claim were true.

Proposition 3 (Standard Motivation)**.** *If markets are efficient, then in every empirical setting as $N \to \infty$ (or, equivalently, under the distribution entailed by the market setting, π) the true model, $\mathcal{A}$, will fit the data at least as well as any alternative model, $\hat{\mathcal{A}} \subseteq \mathcal{K}$:*

$$\mathcal{A} \in \arg\min_{\hat{\mathcal{A}} \subseteq \mathcal{K}} \left\{ \min_{a, b_1, \ldots, b_K} \mathbb{E}\Big[\big(R - a - \textstyle\sum_{k \in \hat{\mathcal{A}}} b_k \cdot X_k \big)^2 \Big] \right\} \tag{16}$$

Unfortunately, while this standard motivation is true, it is also irrelevant to model identification. Model identification involves solving an inverse optimization problem (Kirsch, 2011). The goal is not to find a model that best fists the observed data. This would be the forward problem. Proposition 3 says that investors' true model will always be a solution to this forward problem when markets are efficient.

However, to identify which optimization problem investors are solving, a researcher must verify some unique implication that could only be consistent with the associated asset-pricing model. Proposition 3 guarantees that investors' model will fit the data at least as well as any alternative model. It says nothing about whether investors' model is the only model that fits so well.

In fact, if we take the logic behind Proposition 2 seriously, then it is clear that the assumption of market efficiency makes it impossible for a researcher to identify investors' model. For markets to be efficient, investors must be able to correct every kind of error, cognitive errors included. But, to correct a cognitive error about X, investors must know about at least one correlated predictor. And, because correlated predictors are potential confounds, when a researcher assumes that markets are efficient, he is assuming investors must know about a potential confound for any $\hat{b} > 0$ he estimates. Every empirical setting with a "Yes" in column (4) of Table 1 has a "No" in columns (2) and (3).

Any test of market efficiency is actually a joint test of whether prices are correct and how to calculate prices correctly. Identification is defined relative to the class of models you are willing to consider. This is the Joint Hypothesis Problem Fama (1970, JHP).

It is common practice among asset-pricing researchers to take things a bit further. Under the more extreme interpretation, whenever it looks like investors are in error, the JHP would tell the researcher to assume that investors must know about an omitted variable which would rationalize their behavior. Proposition 2 shows this interpretation is counterproductive. It makes it impossible to identify investors' model.

7 Conclusion

This paper aims to change how asset-pricing researchers approach model identification. The key insight is that, when otherwise intelligent investors fail to correct an error, a researcher learns something about what these investors did not know. They must not have known about anything which would have allowed them to spot their mistake. If they did, then they would have stopped making it. What's more, this logic allows a researcher to learn what does not belong to investors' information set regardless of what belongs in his own information set.

I leverage this insight to show one way that asset-pricing researchers can solve their core omitted-variables problem. If X predicts returns, then any correlated predictor that investors know about but researchers have yet to discover represents a potential confound. I define a special kind of error, called a "cognitive error", which investors will only fail to correct if they are not aware of any such omitted variables. So if X predicts returns and investors fail to correct a cognitive error about X, then a researcher can be sure that investors are pricing assets based on X.

This result turns the existing behavioral-finance literature into a source of instruments for identifying how mostly rational investors price assets. And, to help researchers reframe their thinking, I offer a new metaphor for how a special kind of investor error can be useful for identification purposes: the canary trap.

References

Altonji, J., T. Elder, and C. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*.

Angrist, J. (1990). Lifetime earnings and the Vietnam-era draft lottery: Evidence from Social Security Administrative records. *American Economic Review*.

Angrist, J. and A. Krueger (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*.

Asness, C., A. Frazzini, R. Israel, T. Moskowitz, and L. Pedersen (2018). Size matters, if you control your junk. *Journal of Financial Economics*.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*.

Banerjee, A., S. Chassang, S. Montero, and E. Snowberg (2020). A theory of experimenters: Robustness, randomization, and balance. *American Economic Review*.

Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*.

Barberis, N., A. Shleifer, and J. Wurgler (2005). Comovement. *Journal of Financial Economics*.

Ben-David, I., J. Li, A. Rossi, and Y. Song (2020). Discontinued positive feedback trading and the decline in asset-pricing factor profitability. *Working Paper*.

Berk, J. (1995). A critique of size-related anomalies. *Review of Financial Studies*.

Birru, J. (2015). Confusion of confusions: A test of the disposition effect and momentum. *Review of Financial Studies*.

Boyer, B. (2011). Style-related comovement: Fundamentals or labels? *Journal of Finance*.

Bryzgalova, S. (2017). Spurious factors in linear asset-pricing models. *Working Paper*.

Campbell, J. (2017). *Financial decisions and markets: A course in asset pricing*. Princeton University Press.

Campbell, J. and J. Mei (1993). Where do betas come from? asset-price dynamics and the sources of systematic risk. *Review of Financial Studies*.

Case, K. and R. Shiller (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*.

Chemla, G. and C. Hennessy (2019). Rational expectations and the paradox of policy-relevant natural experiments. *Journal of Monetary Economics*.

Chemla, G. and C. Hennessy (2021a). Equilibrium counterfactuals. *International Economic Review*.

Chemla, G. and C. Hennessy (2021b). Signaling, instrumentation, and cfo decision-making. *Journal of Financial Economics*.

Chen, H., L. Cohen, and U. Gurun (2019). Don't take their word for it: The misclassification of bond mutual funds. *Working Paper*.

Chen, H., W. Dou, and L. Kogan (2021). Measuring "dark matter" in asset-pricing models. *Journal of Finance*.

Chinco, A., S. Hartzmark, and A. Sussman (2020). Necessary evidence for a risk factor's relevance. *Working Paper*.

Clancy, T. (1988). *Patriot games*. Penguin.

Cochrane, J. (2001). *Asset pricing*. Princeton university press.

Cochrane, J. (2011). Presidential address: Discount rates. *Journal of Finance*.

Cochrane, J. (2013). Gene Fama's Nobel Prize. https://johnhcochrane.blogspot.com/2013/10/gene-famas-nobel.html.

Cooper, M., O. Dimitrov, and R. Rau (2001). A `rose.com` by any other name. *Journal of Finance*.

Cooper, M., H. Gulen, and R. Rau (2005). Changing names with style: Mutual fund name changes and their effects on fund flows. *Journal of Finance*.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*.

Deng, Y., J. Quigley, and R. Van Order (2000). Mortgage terminations, heterogeneity, and the exercise of mortgage options. *Econometrica*.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*.

Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics non-parametrically. *Review of Financial Studies*.

Graham, B. and D. Dodd (1934). *Security analysis*. McGraw-Hill New York.

Grinblatt, M. and B. Han (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics*.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*.

Hartzmark, S. and D. Solomon (2020). Reconsidering returns. *Review of Financial Studies*.

Hartzmark, S. and A. Sussman (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *Journal of Finance*.

Harvey, C., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies*.

Heckman, J. (1979). Sample-selection bias as a specification error. *Econometrica*.

Heckman, J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*.

Hennessy, C. and I. Strebulaev (2019). Beyond random assignment: Credible inference and extrapolation in dynamic economies. *Journal of Finance*.

Himmelberg, C., C. Mayer, and T. Sinai (2005). Assessing high house prices: Bubbles, fundamentals and misperceptions. *Journal of Economic Perspectives*.

Hirshleifer, D. (2020). Social-transmission bias in economics and finance. *Journal of Finance*.

Imbens, G. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic literature*.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica*.

Jackson, M. (2010). *Social and economic networks*. Princeton University Press.

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

Kasy, M. (2016). Why experimenters might not always want to randomize, and what they could do instead. *Political Analysis*.

Kirsch, A. (2011). *An introduction to the mathematical theory of inverse problems*. Springer Science & Business Media.

Kozak, S., S. Nagel, and S. Santosh (2018). Shrinking the cross-section. *Journal of Financial Economics*.

Kuchler, T. and J. Stroebel (2020). Social finance. *Working Paper*.

Lucas, R. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester* *conference series on public policy*.

Manski, C. (1995). *Identification problems in the social sciences*. Harvard University Press.

Manski, C. (2009). *Identification for prediction and decision*. Harvard University Press.

Martin, I. and S. Nagel (2019). Market efficiency in the age of big data. *Working Paper*.

McPherson, M., L. Smith-Lovin, and J. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*.

Mehrling, P. (2011). *Fischer Black and the revolutionary idea of finance*. John Wiley & Sons.

Olken, B. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peters, J., P. Bühlmann, and N. Meinshausen (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society: Series B*.

Poterba, J. (1984). Tax subsidies to owner-occupied housing: An asset-market approach. *Quarterly Journal of Economics*.

Rashes, M. (2001). Massively confused investors making conspicuously ignorant choices (MCI–MCIC). *Journal of Finance*.

Ritter, J. (1996). How I helped to make Fischer Black wealthier. *Financial Management*.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*.

Sensoy, B. (2009). Performance evaluation and self-designated benchmark indexes in the mutual-fund industry. *Journal of Financial Economics*.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*.

Shefrin, H. and M. Statman (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*.

Shiller, R. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*.

Shiller, R. and J. Pound (1989). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*.

Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*.

Shleifer, A. (1986). Do demand curves for stocks slope down? *Journal of Finance*.

Thomas, S. (2002). The saga of the first stock-index futures contract: Benchmarks, models, and learning. *Journal of Money, Credit and Banking*.

Tirole, J. (2009). Cognition and incomplete contracts. *American Economic Review*.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science*.

Verma, T. and J. Pearl (1991). Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, 1991*.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*.

A Technical Appendix

Lemma 2 (Verma and Pearl, 1991)**.** *The distributions entailed by two linear structural models are covariance equivalent iff they have the same zero partial correlations.*

Proof. **(Lemma 1)** *When paired with a particular market environment, model m defines a linear structural model given implementation i.*

Thus, we can re-write Definition 6 as saying that an empirical setting is identified for X_1 if it is not possible to create a linear structural model that is covariance equivalent to the one in the current setting by pairing any model $m' \notin \mathcal{M}_1^R(i)$ with any market environment $e \in \mathcal{E}$ under implementation i.

From Lemma 2, we see that this is only possible if investors are not aware of any omitted variable that is correlated with X_1 in this particular empirical setting. □

Proof. **(Proposition 1)** *Suppose investors have failed to correct a cognitive error about X_1 in a researcher's empirical setting. If investors are otherwise intelligent (Assumption ♠), they cannot know about any predictors correlated with X_1 in that setting. Otherwise they would have been able to spot and correct the cognitive error (Equation 9). Thus, the setting is identified.*

When investors make a cognitive error about X_1, they input

$$\tilde{x}_{n,1} = x_{n,1} + \Delta[x_{n,1}] = \sqrt{1-\theta} \cdot x_{n,1} + \sqrt{\theta} \cdot z_n \tag{17}$$

instead of $x_{n,1}$ into their pricing rule.

Hence, when the researcher regresses the cross-section of returns on Z (Equation 10), he will find that

$$\mathrm{Sign}[\hat{b}] = \mathrm{Sign}[\hat{d}] \tag{18}$$

This means that the researcher can draw the same conclusions from $\hat{d}$ as from $\hat{b}$:

$$\hat{d} \begin{cases} > 0 & \Rightarrow \quad \mathcal{M}_1^R(i) = \mathcal{M}^R(i) \\ = 0 & \Rightarrow \quad \mathcal{M}_1^R(i) = \emptyset \end{cases} \tag{19}$$

Thus the identification strategy is a valid identification strategy. □

Proof. **(Proposition 2)** *The researcher only studies empirical settings in which X_1 is uncorrelated with all predictors that he is currently aware of (Assumption B). So, if investors know about some X_k that is correlated with X_1, then it must represent an as-yet-undiscovered omitted-variable in the eyes of the researcher, $k \in \mathcal{K}^I \setminus \mathcal{K}^R$.*

Otherwise intelligent investors can only recognize a cognitive error about X_1 in a particular empirical setting if they know about another correlated predictor X_k in that setting (Equation 9).

Thus, if otherwise intelligent investors can recognize a cognitive error about X_1 in a particular empirical setting, then these investors must know about a correlated omitted variable. So Lemma 2 implies, even if X_1 predicts the cross-section of returns, $\hat{b} > 0$, there must exists an alternative covariance-equivalent model not involving X_1.

Hence, there can be no valid identification strategy in this setting. □

Proof. **(Proposition 3)** *The payout and pricing rules together with distributional assumptions concerning predictors, the nuisance variable, and noise imply a linear structural model with normal payout shocks. In such a model, we have that $-LogLik \propto \mathbb{E}\big[\,(R - a - \sum_{k \in \hat{\mathcal{A}}} b_k \cdot X_k)^2\,\big]$. And the coefficients $\{a, b_1, \ldots, b_K\}$ implied by the data-generating process, by definition, are one way to minimize this negative log likelihood.* □

B Example Economy

This section gives an example economy which would generate an asset-pricing model of the form in Definition 3.

Each risky asset has a supply of $\frac{\psi}{N} > 0$ shares. Assume investors can also invest in a riskless bond, which is elastically supplied and has a zero net return $r_f = 0$.

Let d_n denote the number of shares of the nth risky asset that investors demand. Investors choose how many shares of each risky asset to hold with an eye towards maximizing their expected utility from terminal wealth, W, conditional on the observed predictor and price levels for each asset:

$$\begin{aligned} \underset{d_1,\ldots,d_N}{\text{maximize}} \quad & \mathbb{E}\Big[\,\mathrm{U}(W)\,\Big|\,\{X_1 = x_{n,1}, \ldots, X_K = x_{n,K}, P = p_n\}_{n=1}^{N}\,\Big] \\ \text{subject to} \quad & W = \omega + \textstyle\sum_{n=1}^{N} \{V_n - p_n\} \cdot d_n \end{aligned} \tag{20}$$

$\omega > 0$ is investors' initial wealth, which is constant across environments $e \in \mathcal{E}$.

Investor's model, m, consists of an optimization problem like Equation (20) together with a preference specification, $\mathrm{U}(w)$. The rational-expectations equilibrium (REE) associated with model m consists of a pricing rule $\mathbb{P}^\star(X_1, \ldots, X_K)$ and a demand rule $\mathbb{D}^\star(X_1, \ldots, X_K, P)$ that satisfy the following two conditions:

i) Given the pricing rule $p^\star \leftarrow \mathbb{P}^\star(X_1 = x_1, \ldots, X_K = x_K)$, the demand rule $d^\star \leftarrow \mathbb{D}^\star(X_1 = x_1, \ldots, X_K = x_K, P = p^\star)$ solves the optimization problem in Equation (20) in all market environments $e \in \mathcal{E}$ given investor preferences.

ii) The resulting portfolio allocation clears risky-asset markets, $\frac{\psi}{N} = d_1^\star = \cdots = d_N^\star$.

In this paper, I model an asset-pricing researcher who does not know which predictors are active, $\mathcal{A}$. He is trying to answer the question: Is X_1 is one of these active predictors, $1 \in \mathcal{A}$? That is the source of his uncertainty.

For example, there is disagreement among asset-pricing researchers about whether liquidity risk is a priced risk factor. But there is no disagreement about how exposure to liquidity risk would affect prices if it were part of investors' model.

Since uncertainty is coming from $\mathcal{A}$ not $\mathrm{U}(w)$, let's assume a convenient functional form for $\mathrm{U}(w)$:

$$\mathrm{U}(w) = -\exp\{-\gamma \cdot w\} \quad \text{for some } \gamma > 0 \tag{21}$$

γ is investors' risk aversion, which is the same in every $e \in \mathcal{E}$. Let $\mathcal{M}$ denote the set of all such CARA-utility models defined over environments $\mathcal{E}$.

The REE demand rule associated with model $m \in \mathcal{M}$ will take the following form:

$$\mathbb{D}^{\star}(X_1, \ldots, X_K \,|\, P = p) = \frac{1/\gamma}{\sigma_Y} \cdot \left(\frac{\left\{ \mu_V - \sum_{k \in \mathcal{A}} \phi_k \cdot X_k \right\} - p}{\sigma_Y} \right) \tag{22}$$

Investors will be long, $d^{\star} > 0$, whenever their conditional expectation of an asset's future payout exceeds the price, $\mathbb{E}[V_n \,|\, X_1 = x_1, \ldots, X_K = x_K] = \mu_V - \sum_{k \in \mathcal{A}} \phi_k \cdot x_k > p$. Investors will build a larger position when this difference is larger relative to the conditional payout volatility, σ_Y, or when they are less risk averse.

If the REE pricing rule clears the market for the nth risky asset given this demand rule, then investors must want to hold $\frac{\psi}{N}$ shares of that asset when it is priced at $p_n^{\star}$ dollars per share. We can solve for the REE pricing rule by replacing the left-hand side of Equation (22) with $\frac{\psi}{N}$ and solving for the market-clearing price:

$$\mathbb{P}^{\star}(X_1, \ldots, X_K) = \mu_V - \gamma \cdot \tfrac{\psi}{N} \cdot \sigma_Y^2 - \sum_{k \in \mathcal{A}} \phi_k \cdot X_k \tag{23}$$

The first two terms in Equation (23) denote the average price of a risky asset. $\gamma \cdot \frac{\psi}{N} \cdot \sigma_Y^2$ captures how far an asset's price needs to be reduced below its average payout, μ_V, before risk-averse investors are willing to hold exactly $\frac{\psi}{N}$ shares, each with conditional payout variance of σ_Y^2.

Equation (23) implies a linear pricing rule with coefficients:

$$\lambda_0 = \mu_V - \gamma \cdot \tfrac{\psi}{N} \cdot \sigma_Y^2 \tag{24a}$$

$$\lambda_1 = \phi_1 \tag{24b}$$

$$\vdots$$

$$\lambda_K = \phi_K \tag{24c}$$

This is the functional form specified in Definition 3. Most asset-pricing models (and all factor models) can be fit into this framework. As Cochrane (2001, p146) writes: "Linearity is not restrictive: X_k^2 is just another instrument. The only criticism one can make is that some instrument X_j is important. . . and was omitted."