LOCAL KNOWLEDGE IN FINANCIAL MARKETS

ALEX CHINCO

Abstract. There is a minimum number of transactions, \( N^* \), needed to identify a finite number of shocks via price changes. I refer to this threshold value as the signal recovery bound. When fewer than \( N^* \) transactions have occurred, knowledge about which shocks have taken place is inherently local since traders must use some other information in addition to prices to uncover it—e.g., fundamental analysis or word of mouth. I proceed in 3 steps. First, I show that this signal recovery bound is not only a) readily calculable but also b) independent of the precise details of traders’ cognitive constraints. Second, I embed the signal recovery bound in an information-based asset pricing model à la Kyle (1985) to explore how it constrains would-be arbitrageurs who must wait until \( N^* \) transactions have occurred before drawing the correct inferences from prices. When shocks are sufficiently rare or short-lived, no traders choose to become arbitrageurs. Third, I give examples of how the signal recovery bound applies to a wide variety of common financial settings including residential housing markets, public equity markets, and bond markets.

JEL Classification. D83, G02, G12, G14

Keywords. Compressed Sensing, Bounded Rationality, Local Knowledge, Sparsity

Date: March 10, 2014.
University of Illinois at Urbana-Champaign. alexchinco@gmail.com. (916) 709-9934.
I thank Xavier Gabaix, Aurel Hizmo, Charlie Kahn, Josh Pollet, Vuk Talijan, and Jeff Wurgler for extremely helpful comments and suggestions.
1. Introduction

How many price changes do traders need to see in order to tease out which financial shocks have occurred? I show that there is a well-defined answer to this question that is not only a) readily calculable using techniques from the compressed sensing literature but also b) independent of the precise details of traders’ cognitive constraints. I refer to this minimum number of transactions, $N^\star$, as the signal recovery bound. If fewer than $N^\star$ transactions have occurred, then knowledge about which shocks took place is inherently local since price changes aren’t sufficient to broadcast this information. Traders must use some other information source in addition to price changes—e.g., word of mouth.

1.1. Illustrative Example. It’s easiest to see where the signal recovery bound comes from via a short example. Suppose you moved away from Chicago a year ago, and now you’re moving back and looking for a house. When studying a list of recent sales prices, you find yourself a bit surprised. People seem to have changed their preferences for 1 of 7 different amenities: (1) a 2 car garage, (2) a third bedroom, (3) a half-circle driveway, (4) granite countertops, (5) energy efficient appliances, (6) central A/C, or (7) a walk-in closet? The mystery amenity is raising the sale price of some houses by $\beta > 0$ dollars. You would know exactly how preferences had evolved if you had lived in Chicago the whole time; however, in the absence of this local knowledge, how many sales do you need to see in order to figure out which of the 7 amenities realized the shock?

The answer is 3. How did I arrive at this number? Suppose you found one house with amenities $\{1, 3, 5, 7\}$, a second house with amenities $\{2, 3, 6, 7\}$, and a third house with amenities $\{4, 5, 6, 7\}$. The combination of the price changes for these 3 houses reveals exactly which amenity has been shocked. i.e., if only the first house’s price was too high, $\Delta p_{1,t} = p_{1,t} - E_{t-1}[p_{1,t}] \approx \beta$, then Chicagoans must have changed their preferences for 2 car garages:

\[
\begin{bmatrix}
\Delta p_{1,t} \\
\Delta p_{2,t} \\
\Delta p_{3,t}
\end{bmatrix} = \begin{bmatrix}
\beta + \epsilon_{1,t} \\
\epsilon_{2,t} \\
\epsilon_{3,t}
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix} \begin{bmatrix}
\beta \\
0 \\
\vdots \\
0
\end{bmatrix} + \begin{bmatrix}
\epsilon_{1,t} \\
\epsilon_{2,t} \\
\epsilon_{3,t}
\end{bmatrix} \tag{1}
\]

where $\epsilon_{n,t} \sim N(0, \sigma_\epsilon^2)$ and $\beta \gg \sigma_\epsilon$. By contrast, if $\Delta p_{1,t} \approx \beta$, $\Delta p_{2,t} \approx \beta$, and $\Delta p_{3,t} \approx \beta$, then people must now value walk-in closets much more than they did a year ago.

Here’s the key point. 3 sales is just enough information to answer 7 yes or no questions and rule out the possibility of no change: $7 = 2^3 - 1$. $N = 4$ sales simply narrows your error bars around the exact value of $\beta$. $N = 2$ sales only allows you to distinguish between subsets of amenities. e.g., seeing just the first and second houses with unexpectedly high prices only tells you that people like either half-circle driveways or walk-in closets more. It doesn’t tell
you which one. The problem changes character at $N^*(7, 1) = 3$ observations. When you have seen fewer than $N^* = 3$ sales, information about how preferences have changed is purely local knowledge because prices can’t publicize this news. You must live in Chicago, love ‘da Bears’, and raise your ‘a’s to know it.

To be sure, there are times when it’s entirely obvious what kind of shock took place. e.g., a January 2008 Chicago Tribune article about Priceline.com (PCLN) reported that “a third-quarter earnings surprise sent [the company’s] shares skyward in November, following an earlier announcement that the online travel agency planned to make permanent a no-booking-fees promotion on its airline ticket purchases.”\footnote{DiColo, J. (2008, Jan. 20) Priceline’s Power Looks Promising in Europe, Asia. Chicago Tribune.} No one was confused about why Priceline realized large returns in November 2008. The only problem facing traders was to figure out how much to adjust Priceline’s stock price. Existing information-based asset pricing models are well suited to this setting, and it is not the focus of the current paper. I am interested in the times when traders have to figure out both which shocks occurred and how large they were—e.g., in the residential housing market. It’s here where the signal recovery bound elbows in and makes room for local knowledge.

1.2. Paper Outline. I proceed in 3 steps. First, in Section 2 I extend the illustrative example above by characterizing the signal recovery bound, $N^*(Q, K)$, in a more general setting where assets have $Q$ payout-relevant attributes, only $K$ of these attributes realize a shock, and each asset’s attribute exposures are chosen randomly. Because the signal recovery bound is only relevant when traders are uncertain about which $K$ attributes realized a shock, I study an asymptotically large market where the number of shocks is much smaller than the total number of attributes. Traders must use some rule, $\phi$, to decode the $N$ price changes and identify which subset of $K$ attributes realized a shock. The error rate of the best computationally tractable selection rule goes through a sudden phase change as the number of transactions traders see crosses the signal recovery bound.

What do I mean by the term ‘phase change’? Let’s consider a more realistic example. Suppose that rather than $K = 1$ and $Q = 7$, houses in Chicago have $Q = 400$ attributes and $K = 5$ of them might have changed since you moved away last year. Moreover, suppose that each asset’s sale price is determined by:

$$\Delta p_n = p_n - E[p_n] = \sum_{q=1}^{400} \beta_q \cdot x_{n,q} + \epsilon_n$$

where

$$5 = \|\beta\|_0 = \sum_{q=1}^{400} 1\{\beta_q \neq 0\} \tag{2}$$

with $x_{n,q} \overset{\text{iid}}{\sim} N(0, 1)$, $\epsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2_\epsilon)$, and $\beta \gg \sigma_\epsilon$. There are a number of statistical techniques you could use to identify which 5 attributes realized a shock. e.g., you could use forward...
Phase Change in Selection Rule Error Rate

**Figure 1.** Mean squared error (MSE) of 3 attribute selection rules in a market where each house has \(Q = 400\) attributes and \(K = 5\) of these attributes realize shocks as the number of sales increases from \(N = 15\) to \(N = 400\) using 250 iterations. Left: Home buyers run univariate regressions and keep variables with t-stats exceeding \(\sqrt{2 \cdot \log Q} \approx 3.46\). Middle: Home buyers use same regression procedure, but keep variables with p-values less than \(0.25 \cdot (\hat{K}/Q)\) where \(\hat{K}\) is the number of data-implied parameters in the model. Right: Home buyers select attributes using LASSO. Reads: “All 3 statistical procedures display a sudden drop in their MSE as the redundancy level crosses 77% corresponding to \(N^* \approx 22\) transactions.”

Next, in Section 3 I embed the signal recovery bound in a well known asset pricing model to better flesh out its implications for would be arbitrageurs. Here, I give only a brief overview of the model details for readers familiar with the information-based asset pricing literature. I consider a Kyle (1985)-type model with 3 kinds of agents: asset-specific value

\[
\Delta p_n = \beta_q \cdot x_{n,q} + s_n, \quad \forall q = 1, 2, \ldots, Q
\]
investors, market-wide arbitrageurs, and competitive market makers. Asset values are the sum of attributes-specific shocks. Value investors learn the fundamental value of a single asset. Arbitrageurs begin without any private information, but can deduce which shocks occurred after seeing \( N^* \) prices. Competitive market makers set asset prices equal to their conditional expectation given aggregate demand. When fewer than \( N^* \) sales have taken place, value investors have an informational monopoly, and information about attribute-specific shocks is local. This monopoly power vanishes, arbitrageurs rush in, and prices become more efficient immediately after the number of sales exceeds \( N^* \). No traders choose to become arbitrageurs when shocks are sufficiently rare or short-lived.

Finally, in Section 4 I emphasize the universality of this bound by showing how it applies in many different real world settings. The bound applies to traders in any market where assets have a large number of payout relevant attributes and only a few of attributes matter each period. If a finite number of assets transact each trading period, then prices have a finite bandwidth. What makes the signal recovery bound so general (and so interesting) is that it has nothing to do with arbitrageurs’ cognitive abilities. Given a finite number of price changes, even the most sophisticated arbitrageur can only spot a finite number of shocks to fundamentals. e.g., given 3 house sales you can only distinguish between 7 different shocks and the possibility of no change because 3 prices changes allows for \( 2^3 = 8 \) differences. Giving a variety of examples of how the signal recovery bound constrains traders solidifies this point.

1.3. Related Literature. This paper borrows from and brings together several strands of literature. First, it aims to give a mathematical foundation for F.A. Hayek’s notion of local knowledge as suggested in the title. Indeed, Hayek (1945) gives buying a house as a canonical example of a situation requiring local knowledge. The model formulation relies on the existence of attribute-specific shocks to asset fundamentals. Chinco (2013) provides evidence both that assets realize many different kinds of attribute-specific shocks and also that it is hard for traders to identify which ones are relevant in real time.

The permanent factor structure of returns is no doubt important in the real world; however, it’s already been extensively studied. Campbell, Lettau, Malkiel, and Xu (2001) give evidence that the usual factor models only account for a fraction of firm-specific return volatility. e.g., if you selected an NYSE/AMEX/NASDAQ stock at random in 1999, market and industry factors only accounted for 30% of the variation in its daily returns. Recent work by Ang, Hodrick, Xing, and Zhang (2006), Chen and Petkova (2012), and Herskovic, Kelly, Lustig, and Van Nieuwerburgh (2014) gives strong evidence that there is a lot of cross-sectional structure in the remaining 70% of so-called idiosyncratic volatility. i.e., patterns in past idiosyncratic volatility are strong predictors of future returns. Thus, some portion of the 70% remainder appears to be neither permanent factor exposure nor fully idiosyncratic events.
The paper is also closely related to the literature on bounded rationality; yet, there is a fundamental difference in approaches. Existing theories use cognitive constraints to induce boundedly rational decision making. e.g., papers like Sims (2006) and Hong, Stein, and Yu (2007) suggest that cognitive costs force traders use overly simplified mental models, and Gabaix (2011) derives the sort of mental models that traders would choose when facing \( \ell_1 \) thinking costs. By contrast, I use bandwidth constraints on prices to generate similar behavior. Both channels are at work in assets markets. This paper is the first to articulate the bandwidth constraint. The compressed sensing literature originated with Candes and Tao (2005) and Donoho (2006). See Candes and Wakin (2008) for an overview of compressed sensing.

2. Signal Recovery Bound

This section characterizes the signal recovery bound, \( N^*(Q, K) \), when assets have \( Q \) payout-relevant attributes, only \( K \ll Q \) of these attributes realize a shock, and each asset’s attribute exposure is chosen at random. While I motivate the bound with an application to the residential housing market, it is much more general. It applies to traders in any market where assets have a large number of payout relevant attributes and only a few of attributes matter. I begin in Subsection 2.1 by describing traders’ inference problem. Traders must use some rule, \( \phi \), to decode \( N \) price changes and identify which subset of \( K \) attributes realized a shock. In Subsection 2.2 I then show that the error rate of the best computationally tractable rule goes through a sudden phase change as the number of transactions traders see crosses the bound. Finally, in Subsection 2.3 I discuss the signal recovery bound’s interpretation.

2.1. Inference Problem. This subsection describes traders’ inference problem. I study a market with \( N \) assets. Each asset has exposure to \( Q \) different attributes and only \( K \ll Q \) of these attributes realizes a shock. e.g., if the assets are houses, the attributes might be things like having a 2 car garage or granite countertops. Alternatively, if the assets are stocks, the attributes might represent each company’s headquarter location, major shareholders, etc... Traders see price changes of \( \Delta p_{n,t} \) in units of dollars:

\[
\Delta p_{n,t} = p_{n,t} - \mathbb{E}_{t-1}[p_{n,t}] = \sum_{q=1}^{Q} \beta_{q,t} \cdot x_{n,q} + \epsilon_{n,t} \tag{4}
\]

where \( \beta_{n,t} \in \{0, 1/\sqrt{\pi}\} \) denotes a shock to the \( q \)th attribute at time \( t \) in units of dollars per attribute, \( x_{n,q} \sim \mathcal{N}(0, 1) \) denotes the extent to which asset \( n \) displays the \( q \)th attribute, and \( \epsilon_{n,t} \sim \mathcal{N}(0, \sigma_{\epsilon}^2) \) denotes idiosyncratic shocks at time \( t \) that only affect asset \( n \) in units of dollars. Attribute exposures can be thought of as answers to yes/no questions and are thus dimensionless.

Importantly, while the attribute exposures are carefully selected in the illustrative example
in Subsection 1.1, in the remainder of the paper I consider the case where the exposures are randomly assigned. If the assets are houses, then you can think about each $\Delta p_{n,t}$ as a deviation in the sale price from their expected values. If the assets are stocks, then you can think about each $\Delta p_{n,t}$ as an abnormal return with the permanent factor exposures baked into the expectation operator. I drop the time subscripts wherever it causes no confusion.

Traders’ want to figure out which subset $\mathcal{K} \subseteq \mathcal{Q}$ of the attributes has realized a shock:

$$\mathcal{K} = \{ q \in \{1,2,\ldots,Q\} \mid \beta_q \neq 0 \}$$

by studying the $N$ realized price changes $\Delta p_1, \Delta p_2, \ldots, \Delta p_N$. I use the convention that $|\mathcal{K}| = K$ denotes the size of a set. The choice of $\beta_q = 1/\sqrt{K}$ for all $q \in \mathcal{K}$ ensures that the amount of fundamental volatility in the market stays constant as the number of shocks expands and contracts with $\text{Var}[\Delta p_n | \epsilon_n] = \$1$ for any choice of $K$ and $Q$. It is an analytical convenience and not crucial to the results. e.g., all of the analysis goes through if the size of each of the $K$ shocks is a positive random variable.

To capture the idea that the $K$ different shocks are hard to identify, I study the regime where all 3 variables $K, N, Q \to \infty$ with $K/Q \to 0$. e.g., you might take $K = \sqrt{Q}$. Only a vanishingly small fraction of the possible payout relevant attributes realize a shock each period. The scale of financial markets matters. e.g., Daniel (2009) notes that during the Quant Meltdown of August 2007 “markets appeared calm to non-quantitative investors...you could not tell that anything was happening without quant goggles” even though large funds like Highbridge Capital Management were suffering losses on the order of 16% of assets under management.\footnote{Zuckerman, G., J. Hagerty, and D. Gauthier-Villars (2007, Aug. 10) Impact of Mortgage Crisis Spreads. The Wall Street Journal.} All stocks held in long/short market neutral strategies realized a large negative price shock, but this shock was just one of many plausible shocks that could hit equity markets each trading period. Thus, unless you knew where to look (i.e., were wearing “quant goggles”) it passed you right by. Like stage magic, these events took place before your very eyes while you were looking elsewhere.

Traders are looking for a selection rule $\phi(\Delta p, X)$ which eats an $(N \times 1)$-dimensional vector of price changes as well as a $(N \times Q)$-dimensional matrix of attribute exposures and spits out the list of shocked attributes:

$$\phi : \mathbb{R}^N \times \mathbb{R}^{N \times Q} \mapsto \{ \mathcal{J} \subseteq \mathcal{Q} : |\mathcal{J}| = K \}$$

I define the error rate of a selection rule $\phi$ as:

$$\text{Err}[\phi] = \frac{1}{(\frac{Q}{K})} \cdot \sum_{\substack{\mathcal{K} \in \mathcal{Q} \\ |\mathcal{K}| = K}} \text{Pr} \left[ \phi(y, X) \neq \mathcal{K} \mid \mathcal{K} \right]$$

In words, $\text{Pr}[\phi(y, X) \neq \mathcal{K} | \mathcal{K}]$ denotes the probability that the selection rule $\phi$ chooses the
wrong subset of attributes (i.e., makes an error) given the true support \( K \) and averaging over not only the measurement noise, \( \epsilon \), but also the choice of the Gaussian attribute exposure matrix, \( X \). The error rate operator, \( \text{Err}[\cdot] \), then computes a weighted average of these probabilities over every shock set of size \( K \). This is analogous to the error rate measure reported on the \( y \)-axis of Figure 1.

The only cognitive constraint that traders face is that their selection rule must be computationally tractable—i.e., that \( \phi \) can be implemented as a convex optimization program. Under minimal assumptions a convex optimization program is computationally tractable in the sense that the computational effort required to solve the problem to a given accuracy grows moderately with the dimensions of the problem. Natarajan (1995) explicitly shows that \( \ell_0 \) constrained linear programming is NP-hard. This cognitive constraint is really weak in the sense that any selection rule that you might look up in an econometrics or statistics textbook (e.g., forward stepwise regression or LASSO) is going to be computationally tractable. After all, they have to be executed on computers.

2.2. Phase Change. This subsection shows that the error rate of the best computationally tractable rule goes through a sudden phase change as the number of transactions traders see crosses the signal recovery bound. I am interested in whether or not a particular selection rule can decode price changes and identify the shocked attributes in an asymptotically large market. I encode this idea as the notion of asymptotic reliability. A selection rule, \( \phi \), is asymptotically reliable if:

\[
\lim_{K,N,Q \to \infty} \frac{K}{Q} \to 0 \quad \text{and} \quad \beta = \frac{1}{\sqrt{K}} \quad \text{then:}
\]

\[
\lim_{K,N,Q \to \infty} \frac{K}{Q} \to 0 \quad \text{and} \quad \beta = \frac{1}{\sqrt{K}} \quad \text{then:}
\]

\[
\lim_{K,N,Q \to \infty} \text{Err}[\phi] = 0
\]

whereas it is asymptotically unreliable if there exists some constant \( A > 0 \) such that \( \text{Err}[\phi] \geq A \) as \( K, N, Q \to \infty \) with \( K/Q \to 0 \). Proposition 1 below characterizes how the asymptotic reliability of the best possible selection rule changes as I increase the number of transactions.

**Proposition 1** (Phase Change). If \( K, N, Q \to \infty, \frac{K}{Q} \to 0, (N - K) \cdot \beta \to \infty, \) and \( \beta = \frac{1}{\sqrt{K}}, \) then:

1. **There exists an asymptotically reliable selection rule \( \phi \) if for some constant \( a > 0 \):**

\[
N > a \cdot K \cdot \log(Q/K)
\]

2. **No asymptotically reliable selection rule \( \phi \) exists if for some constant \( a' \in (0, a) \):**

\[
N < a' \cdot K \cdot \log(Q/K)
\]

Thus, there exists a signal recovery bound, \( N^*(Q, K) = \Theta(K \cdot \log(Q/K)) \), separating the regions where asymptotically reliable selection is possible and impossible.
This result builds primarily on analysis done in Donoho and Tanner (2009) and Wainwright (2009). To get a sense for what this result means, note the tight agreement between i) where Proposition 1 says the phase change should occur in the asymptotic case and ii) where the phase change actually occurs in Figure 1 in the finite case with $Q = 400$ attributes and $K = 5$ shocks. In each of the 3 panels, the error rate drops to almost 0 at some level of data redundancy between 78% (FDR Threshold, middle) and 84% (LASSO, right). This range of data redundancies corresponds to a data-implied signal recovery bound of somewhere below $\hat{N}^* = 22.73$ transactions since:

$$0.78 = \frac{22.73 - 5}{22.73}$$

(11)

Thus, the empirically derived value, $\hat{N}^*$, almost exactly achieves the theoretically implied value given by Proposition 1:

$$N^*(400, 5) = 5 \cdot \log(400/5) = 21.91$$

(12)

This back of the envelope calculation shows both that popular statistical procedures come close to achieve optimal recovery as soon as possible and that this asymptotic result holds in real world applications.

2.3. Interpreting the Bound. This subsection discusses 4 nuances of interpreting and applying the signal recovery bound described in Proposition 1 above. First, it’s important that each asset’s attribute exposures are drawn iid, $x_{n,q} \iid N(0, 1)$, though the use of the Gaussian distribution is purely for analytical convenience. It’s useful here to draw on an analogy to randomized control trials. i.e., randomizing which assets get sold makes price changes more informative about shocks to fundamentals in the same way that randomizing which subjects get treated in a medical study makes the experimental results more informative about the effectiveness of the drug in question. Why exactly does randomization help? Suppose all of the people who got the real drug recovered and all of the people who got the placebo didn’t. Randomly assigning patients to the treatment and control groups makes it exceptionally unlikely that the patients who took the real drug will happen to have some other trait (e.g., a genetic variation) that actually explains their recovery. A social planner trying to maximize the informativeness of asset prices would push random assets to the market for the same reason. Randomizing attributes decreases the probability that 2 different $K$-sparse vectors $\beta$ and $\beta'$ in Equation (4) are observationally equivalent. Allowing for correlated attributes increases the signal recovery bound, but does not eliminate the phase change.

Second, the result in Proposition 1 holds in the particular asymptotic case where $K/Q \to 0$. There is a different limiting behavior when the number of shocks grows linearly with the number of attributes, $K/Q \to \kappa > 0$, which is described in Wainwright (2009). Regardless of the chosen asymptotics, there is a qualitative change in the nature of any inference problem when
you move from choosing which variables matter to deciding how much the chosen variables matter. To illustrate, consider a short example: $\beta$ usually isn’t sparse in econometric textbooks. When every one of the $Q$ attributes matters, it’s easy to decide which attributes to pay attention to—i.e., all of them. Nevertheless, there is a similar signal recovery bound in the non-sparse case corresponding to the usual $N \geq Q$ requirement for identification. To see why, let’s return to the motivating example in Subsection 1.1, and consider the case where any of the 7 attributes could have realized a shock. This leaves us with 128 different shock combinations:

$$128 = \sum_{k=0}^{7} \binom{7}{k}$$

$$= 1 + 7 + 21 + 35 + 35 + 21 + 7 + 1$$

$$= 2^7$$

so that $N^* = 7$ gives just enough differences to identify which combination of shocks was realized. More generally, we have that for any number of attributes, $Q$:

$$2^Q = \sum_{k=0}^{Q} \binom{Q}{k}$$

This gives an interesting information theoretic interpretation to the meaning of “just identified” that has nothing to do with linear algebra or the invertibility of a matrix.

Third, consider the requirement that an asymptotically reliable selection rule be tractable—i.e., able to be implemented as a convex program. How much harder could the non-convex approach be? A lot. Suppose I told you exactly which $K = 5$ of the attributes had realized a shock. For this sub-problem you could easily estimate the values of each of the coefficients using a standard regression procedure—i.e., a convex approach. You can certainly solve the general problem by solving each of the $\binom{400}{5} \approx 8.3 \times 10^{10}$ sub-problems; however, this is a huge number of cases to check on par with the number of bits in the human genome. As Rockafellar (1993) writes, “the great watershed in optimization isn’t between linearity and nonlinearity, but convexity and non-convexity.” In order to break the signal recovery bound, you would need additional knowledge which eliminates many of these sub-problems from contention. e.g., in the introductory example in Subsection 1.1 you might need to have frequent discussions with your neighbors that dramatically narrow the list of possible preference shocks which might have taken place.

Finally, real world markets are large but finite. By contrast, I derive the signal recovery bound in an asymptotic setting where the numbers of assets, attributes, and shocks all tend toward infinity. At first, this might seem to pose a problem when applying the bound; however, in practice this turns out not to be the case for 2 reasons. First, although real world markets are finite, they are really, really large. Thus, the asymptotic approximation is
a good one. I give more evidence in favor of this starting point in Section 4 below. Second, while it isn’t possible to give a precise formulation of the signal recovery bound in the finite sample case, practical compressed sensing techniques such as the Dantzig Selector introduced in Candès and Tao (2007) can make error rate guarantees in the finite case which are close to the asymptotic results given in the current paper. We regularly make this sort of asymptotic-to-finite leaps in mainstream econometric applications. e.g., practical application of GMM involves a 2-step procedure. The first step estimates the coefficient vector using the identity weighting matrix on the basis that any positive semidefinite weighting matrix will give the same point estimates in the large $T$ limit, and the second step then uses the realized point estimates to compute the optimal weighting matrix and coefficient standard errors. Thus, the use of the signal recovery bound in finite applications seems no more or less extraordinary than other uses asymptotic statistical results in finite econometric applications.

3. Equilibrium Model

This section couches the signal recovery bound described above in a Kyle (1985)-type equilibrium model to explore its consequences for would-be arbitrageurs. First, in Subsection 3.1 I outline the market structure and describe each agent’s problem. Even though the signal recovery bound is a constraint on the maximum bandwidth of a finite number of prices, it nevertheless behaves a lot like a constraint on arbitrageurs’ cognitive abilities. If a shock hasn’t yet manifested itself in $N^*$ transactions, then arbitrageurs can’t discover it by studying only prices. Then, in Subsection 3.2 I characterize the resulting equilibrium and analyze its properties. Finally, in Subsection 3.3 I endogenize each trader’s decision to become either an asset-specific value investor or a market-wide arbitrageur and show that no traders choose to become arbitrageurs when shocks are sufficiently rare or short-lived.

3.1. Market Structure. This subsection outlines the market structure and describes each agent’s problem. I study a Kyle (1985)-type market with $N$ assets whose fundamental values, $v_n$, are governed by their exposure to $Q$ different payout relevant attributes. $Q$ denotes the set of all attributes, and $K \subset Q$ denotes the subset of $K$ attributes which realized a shock of size $\beta_q$. I write the value of each asset $n$ as:

$$v_n = \sum_{q=1}^{Q} \beta_q \cdot x_{n,q} \quad \text{with} \quad x_{n,q} \overset{iid}{\sim} \mathcal{N}(0, 1), \quad \beta_q = \begin{cases} 1/\sqrt{K} & \text{if } q \in K \\ 0 & \text{else} \end{cases}$$

(15)

where $x_{n,q}$ denotes asset $n$’s exposure to the $q$th attribute. Each asset’s fundamental value, $v_n$, has units of dollars, and each shock, $\beta_q$, has units of dollars per attribute.

There are 2 kinds of optimizing agents, asset-specific value investors and market-wide arbitrageurs, as well as a collection of competitive market makers. Figure 2 describes the
Timing in Baseline Model

Figure 2. What each agent knows and when they know it in the baseline model. Reads: “At time $t = N^*$ market-wide arbitrageurs reach the critical number of observations necessary to infer which attribute-specific shocks occurred and begin to trade like asset-specific value investors in all assets traded at times $t \geq N^* + 1$.”

timing of the model. The model starts in period $t = 0$ with nature assigning attribute exposures to each of the $N$ assets and picking a subset of $K$ attributes, $K \subset Q$, to realize shocks. After the shocks and the attribute exposures have been drawn, each of the $N$ asset-specific value investors learn the true value of their asset, $v_n$. Asset-specific value investors immediately learn the true value of only 1 asset. By contrast, market-wide arbitrageurs can deduce the fundamental value of all $N$ assets from prices, but must wait until $N^*$ trades have already taken place to so. This is the key implication of the signal recovery bound.

In periods $t = 1, 2, \ldots, N$ each of the $N$ assets sequentially comes to market, and all agents are aware of the number of assets that have previously traded. Importantly, everyone knows if the $N^*$ transactions threshold has been crossed. Asset-specific value investors solve the standard static Kyle (1985) optimization problem with risk neutral preferences given their knowledge of $v_n$:

$$
\pi^{VI}_n = \max_{y^{VI}_n} E^{VI}_n \left[ (v_n - p_n) \cdot y^{VI}_n \right]
$$

where $y^{VI}_n$ denotes the size of the market order that asset $n$’s value investors place in units of shares. Each value investor’s problem is static because every asset that comes to market has a different set of attribute exposures. Market-wide arbitrageurs solve a similar problem once the $N^*$ transaction threshold has been reached:

$$
\pi^{Arb} = \max_{\{y^{Arb}_{n=N^*+1}\}} \sum_{n=N^*+1}^N E^{Arb}_n \left[ (v_n - p_n) \cdot y^{Arb}_n \right] - c \cdot N^*
$$

where $y^{Arb}_n$ likewise denotes the size of the market order that arbitrageurs place for $n$ in units of shares. However, for each period that they wait, arbitrageurs have to pay a cost of $c \geq 0$ dollars. If $c = 0$, then the only cost born by inactive arbitrageurs is the lost profits from not
trading; whereas, if $c > 0$, then arbitrageurs face additional inactivity costs. e.g., you can think of this penalty as an opportunity cost of cash or increases in transactions costs due to lower trading volume.

Market makers in each of the $N$ assets observe the aggregate order flow:

$$y_n = y_n^{VI} + y_n^{Arb} + \xi_n \quad \text{with} \quad \xi_n \overset{\text{iid}}{\sim} N(0, \sigma_\xi^2)$$

(18)

which is composed of demand from the asset-specific value investors, market-wide arbitrageurs, and noise traders, $\xi_n$. Perfect competition implies that market makers set each asset’s price equal to the conditional expectation of its fundamental value given the realized aggregate order flow:

$$p_n = E[v_n | y_n]$$

(19)

Importantly, market makers do not immediately know which shocks have occurred once $N^*$ transactions have taken place. Instead, it takes until the $N$th asset is sold for market makers to learn the truth. The idea here is that arbitrageurs aren’t superhuman, but they do have skill. i.e., they can’t solve an NP-hard non-convex optimization problem, but they are very good at interpreting price signals. In fact, in the model they are the only ones talented enough to actually achieve the signal recovery bound and spot which shocks have occurred after only $N^*$ transactions have taken place. I interpret the total number of assets that come to market, $N$, as duration of the attribute-specific shocks.

3.2. Baseline Equilibrium. This subsection characterizes the resulting equilibrium in this market. The equilibrium concept is standard for noisy rational expectations models. An equilibrium is a linear demand rule for each group of asset-specific value investors as well as the market-wide arbitrageurs:

$$y_n^{VI} = v_n \times \begin{cases} \theta^{VI} & \text{if } n > N^* \\ \theta^{VI}_0 & \text{else} \end{cases} \quad \text{and} \quad y_n^{Arb} = v_n \times \begin{cases} \theta^{Arb} & \text{if } n > N^* \\ 0 & \text{else} \end{cases}$$

(20)

and a linear pricing rule for the market makers:

$$p_n = y_n \times \begin{cases} \lambda_* & \text{if } n > N^* \\ \lambda_0 & \text{else} \end{cases}$$

(21)

such that: (a) given the pricing rule the demand rules solve the optimization programs in Equations (16) and (17), and (b) given the demand rules, the pricing rule satisfies the boundary condition in Equation (19).

How does this setup map onto the inference problem described in Section 2 above? First, the case-wise linear pricing and demand rules stem from the phase change at $N^*$. When $N^*$ or fewer transactions have occurred, market-wide arbitrageurs are less informed than the
market makers so they do not enter. After $N^*$ transactions have occurred, they know the true $\beta$ and are thus equally informed as the asset-specific value investors. As a result, arbitrageurs draw their inferences using only the coefficients $\theta_o$ and $\lambda_o$. Since the unconditional mean of each asset’s value is $E[v_n] = 0$, it’s clear that $\Delta p_n = p_n - E[p_n] = p_n$. Thus, for market-wide arbitrageurs trying to deduce which attribute-specific shocks occurred, the relevant inference problem is given by:

$$\Delta p_n = \lambda_o \cdot \left( \theta_o \cdot \sum_{q=1}^{Q} \frac{1_{(q \in K)}}{\sqrt{K}} \cdot x_{n,q} + \xi_n \right)$$  \hspace{1cm} (22)

The proposition below characterizes the equilibrium coefficients subject to market-wide arbitrageurs’ inability to solve this inference problem before seeing at least $N^*$ transactions.

**Proposition 2** (Baseline Equilibrium). In the limit as $N, Q, K \to \infty$ and $Q/K \to 0$, the equilibrium price impact and demand coefficients are given by:

$$\lambda_o = \frac{1}{2\sigma_\xi}, \quad \theta_o = \sigma_\xi, \quad \lambda_* = \frac{\sqrt{2}}{3\sigma_\xi}, \quad \theta_* = \frac{\sigma_\xi}{\sqrt{2}}$$  \hspace{1cm} (23)

where $(\lambda_o, \theta_o)$ denote the prevailing coefficients when no more than $N^*$ transactions have taken place and $(\lambda_*, \theta_*)$ denote the prevailing coefficients when more than $N^*$ transactions have occurred.

Note the sharp change in the informativeness of prices as the number of transactions crosses the $N^*(Q, K)$ threshold. When fewer than $N^*$ sales have taken place, value investors have an informational monopoly, and information about attribute-specific shocks is local. This monopoly power vanishes, arbitrageurs rush in, and prices become more efficient immediately after the number of sales exceeds $N^*$. To gauge the information content of prices, I compute how much an asset’s price moves with its fundamental per unit of variation in its fundamental value, $\frac{\text{Cov}[v_n, p_n]}{\text{Var}[v_n]}$, in units of $1/\text{assets}$. When this measure is close to 1, then prices reflect most of the realized fundamentals; whereas, when this measure is close to 0, prices are reveal very little information about the realized fundamentals. A simple computation shows that when only value investors are in the market:

$$\frac{\text{Cov}_o[v_n, p_n]}{\text{Var}[v_n]} = \text{Cov} [v_n, p_n | n \leq N^*] = \frac{1}{2}$$ \hspace{1cm} (24)

By contrast, this statistic grows by 33% when value investors’ informational monopoly vanishes and arbitrageurs enter the market:

$$\frac{\text{Cov}_*[v_n, p_n]}{\text{Var}[v_n]} = \text{Cov} [v_n, p_n | n > N^*] = \frac{2}{3}$$  \hspace{1cm} (25)

What’s new here? The actual functional forms of the price impact and demand coefficients are quite standard in Proposition 2. Indeed, Holden and Subrahmanyam (1992) shows that
imperfect competition among informed traders with the same private information leads to aggressive competition and rapid revelation of their signal. The novel feature is the link between which set of coefficients is prevailing and the scale of the market. e.g., earlier work has shown that imperfect competition among informed traders can lead to more informative prices, but when should we expect to see this imperfect competition in markets? The signal recovery bound characterized above gives a precise quantitative statement about when even the most sophisticated market-wide arbitrageurs can enter. This is where things get interesting. We already knew that prices should be different in a market with competition among traders, this theory couches results from the compressed sensing literature in a well-known economic model to make predictions about when this competition can and can’t emerge.

Note that while the volatility of noise trader demand alters the price impact and demand coefficients, it does not alter the threshold number of transactions, $N^*$, at which market-wide arbitrageurs can deduce which attribute-specific shocks took place. This is a consequence of studying the limiting case as the numbers of attributes, shocks, and assets all tend toward infinity. If $Q$, $K$, and $N$ are all finite as in Figure 1, then there will be times when market-wide arbitrageurs will be able to deduce a few (rather than all) of the shocks. For reasons discussed above this impact will be second order; moreover, numerical investigations such as those in Section 2 show that popular statistical procedures nearly achieve the signal recovery bound in finite samples in the presence of noise.

3.3. Learning Style Choice. This subsection endogenizes traders’ decision to become either an asset-specific value investor or a market-wide arbitrageur. After all, traders don’t just randomly decide how they learn about asset fundamentals. This is a strategic decision. When shocks are rarer or shorter-lived compared to the number of transactions each period, it will be relatively more profitable to be a value-investor. In such a world, market-wide arbitrageurs will only be able to trade a few assets, $N - N^*$, after waiting $N^*$ transactions to draw their inference. As a result, it will be more likely at a value investor chosen will be an informational monopolist.

I link the number of assets market-wide arbitrageurs can trade to the number of transactions required by the signal recovery bound via:

$$\Delta N = N - N^* = \eta \cdot N^*$$

(26)

where I refer to the dimensionless parameter $\eta$ as the coverage ratio. Roughly speaking, the coverage ratio captures how awesome it would be to be the only market-wide arbitrageur. When the coverage ratio is large, $\eta \gg 0$, attribute-specific shocks are not very rare compared to the number of transactions per period. In this world arbitrageurs can quickly identify which shocks occurred via price changes, and then have a large number of subsequent periods in which to exploit this information. When the coverage ratio is a small positive number,
Timing in Model with Learning Style Choice

\[ \eta \approx 0^+, \] then attribute-specific shocks are rare compared to the number of transactions per period. It is possible for market-wide arbitrageurs to uncover which attributes realized a shock by only looking at prices; however, after making this discovery, they will have very few periods in which to exploit this info. Finally, if the coverage ratio is negative, \( \eta < 0 \), then it is impossible for a market-wide arbitrageur to back out which attribute-specific shocks occurred by using nothing but prices.

In a model where agents make a learning style choice, it is now important to keep track of the number of agents. I consider a market with a unit mass of ex ante identical traders. Let \( \omega \in (0,1] \) denote the fraction of traders that decide to become value investors in some asset. As illustrated in Figure 3, the structure of the model where agents face a learning style choice is identical to that of the baseline model except for the fact that there is an additional period tacked onto the beginning of time where each agent chooses whether to become an asset-specific value investor or a market-wide arbitrageur.

The equilibrium concept generalizes the one introduced above in Subsection 3.2. An equilibrium is a linear demand rule for each group of asset-specific value investors as well as the market-wide arbitrageurs, a linear pricing rule for the market makers, and a learning style choice \( \omega \) such that: (a) given the pricing rule the demand rules solve the optimization programs in Equations (16) and (17), (b) given the demand rules, the pricing rule satisfies the boundary condition in Equation (19), and (c) given the pricing rule and demand rules no trader wants to unilaterally change from being a value investor to an arbitrageur or vice versa.
Equilibrium with Endogenous Learning Style

Figure 4. Equilibrium where each trader can choose to become either an asset-specific value investor or a market-wide arbitrageur when arbitrageur’s opportunity cost is \( c = 0.10 \) in units of \$/asset. The x-axis denotes the asset coverage ratio and is increasing in the total number of traded assets relative to the signal recovery bound. The left-most panel reports the fraction of traders that choose to become value investors. The middle panel reports the average information content of prices in units of \( 1/\text{assets} \) where the average is computed across assets. The right-most panel reports the average profit per asset earned by traders in units of \$/asset where the average is computed across traders. Reads: “As you continue to move from left to right following the green dashed line (\( \sigma_\xi = 1.0 \)), you reach a point at \( \eta = 0.44 \) where becoming a market-wide arbitrageur suddenly becomes worthwhile. At this point, the first arbitrageur rushes in and not only dramatically increases the informativeness of prices but also radically decreases the profit per asset earned by each trader.”

The proposition below solves for the value of \( \omega \) that makes the marginal trader indifferent between becoming a asset-specific value investor in some asset and being a market-wide arbitrageur.

**Proposition 3** (Endogenous Learning Style). If arbitrageurs’ opportunity cost of delay satisfies \( c < \frac{\eta \cdot \sigma_\xi}{3 \cdot \sqrt{2}} \), then in the limit as \( N, Q, K \to \infty \), \( Q/K \to 0 \), and \( \Delta N/N^* \to \eta \) the expected profit earned by asset-specific value investors and market-wide arbitrageurs is given by:

\[
E[\pi_{VI}^n] = \sigma_\xi \cdot \left( \frac{3 + \sqrt{2} \cdot \eta}{6(1+\eta)} \right) \\
E[\pi_{Arb}] = \left( \frac{\eta \cdot \sigma_\xi - 3 \cdot \sqrt{2} \cdot c}{3 \cdot \sqrt{2}} \right) \cdot N^*
\]

and the fraction of traders that decide to be value investors given by:

\[
\omega = \min \left[ 1, \frac{\sigma_\xi}{3} \cdot \left( \frac{3 \cdot \sqrt{2} + 2 \cdot \eta}{\sqrt{2} \cdot (\sigma_\xi - c) + \sigma_\xi \cdot \eta} \right) \right]
\]

Figure 4 gives a sense of the equilibrium properties. First, consider how the equilibrium changes as the coverage ratio increases—i.e., moving from left to right. When the coverage ratio is small, \( \eta \approx 0^+ \), no traders choose to become arbitrageurs as shown in the left-most panel. In this world, attribute-specific shocks are too rare and short-lived. After waiting around for \( N^* \) transactions to take place at an opportunity cost of \( c \) dollars per transaction, arbitrageurs have too few subsequent chances, \( \Delta N \), to exploit any information they uncover.
As the coverage ratio rises, nothing happens at first—i.e., there is a region where the fraction of value investors holds constant at $\omega = 1$. However, as you continue to move from left to right you reach a point where becoming a market-wide arbitrageur suddenly becomes worthwhile. At this point, the first arbitrageur rushes in and dramatically affects prices. e.g., the middle panel shows that the information content of prices discontinuously lurches upward when market-wide arbitrageurs enter the scene. If prices are more informative, then traders must be worse off. This is exactly what the right-most panel shows. The presence of market-wide arbitrageurs leads to competition among traders and thus lower profits per trader. The traders would all be better off if they could commit to remaining asset-specific value investors; however, this cannot be an equilibrium because the first trader to deviate and learn from the cross-section of past prices would earn huge profits.

Next, consider how the equilibrium changes as the noise trader demand volatility increases—i.e., comparing the solid red ($\sigma_\xi = 0.6$) to the dashed green ($\sigma_\xi = 1.0$) lines. Increasing noise trader demand volatility makes it harder for market makers to distinguish between informed trades and noise trader demand and thus grows the potential profits available to traders. e.g., you can see this in the right-most panel where the dashed green ($\sigma_\xi = 1.0$) is well above the solid red line ($\sigma_\xi = 0.6$) in the regime where $\eta \approx 0^+$ and all traders are value investors. However, increasing noise trader demand volatility also makes arbitrageurs enter the market sooner. e.g., in the left-most panel it’s clear that the first trader decides to become a market-wide arbitrageur at $\eta = 0.44$ when $\sigma_\xi = 1.0$; whereas, no traders decide to become a market-wide arbitrageur until $\eta = 0.71$ when $\sigma_\xi = 0.6$. An interesting consequence of this tradeoff is that for moderate coverage ratios, increasing noise trader demand volatility can actually make prices more informative.

This section highlights the broad array of applications to which the signal recovery bound applies. First, in Subsection 4.1 I study the case of an company analyst trying to figure out how aggressively to trade on private information. This example illustrates how the signal recovery bound acts as if it was a constraint on would be arbitrageurs’ cognitive abilities. Next, in Subsection 4.2 I examine the conditions under which it’s worth it for a trader to hunt down the root cause of an arbitrage opportunity in the Treasury market. Finally, in Subsection 4.3 I ask the question: “Is it possible to create a price index for a single house?”

4.1. Analyst’s Advantage. This subsection illustrates how the signal recovery bound acts as if it was a constraint on would be arbitrageurs’ cognitive abilities. Imagine you’re a value investor studying Baker Hughes (BHI), a company in the petroleum equipment and services industry. Through caffeine-fueled nights of research you’ve discovered that the company is
WSJ Articles with “Petroleum Industry” as Subject

Figure 5. Petroleum equipment and services industry article descriptive tags in ProQuest. Reads: “If you select a Wall Street Journal article on the petroleum equipment and services industry over the period from 2011 to 2013 there is a 19% chance that ‘Oil sands’ is a listed descriptor and a 7% chance that ‘LNG’ (i.e., liquid natural gas) is a listed descriptor.”

due for a big unexpected payout. i.e., in the language of Section 3 you know that:

\[ v_{\text{BHI}} \gg 0 \]  

(29)

Your private signal is really valuable information about Baker Hughes’ value. How aggressively should you trade on it? In the canonical Kyle (1985)-type setup, you face the following trade off. On one hand, you want to build up a large position in Baker Hughes to take advantage of the big payout that you know is going to happen in the future. On the other hand, you don’t want to trade too aggressively and allow uninformed traders to back out your private signal about Baker Hughes from prices.

This paper points out an additional consideration: whether or not would be arbitrageurs can back out why the shock to Baker Hughes occurred from prices. e.g., suppose Baker Hughes’ value shock is due to positive news for all firms who’ve heavily invested in hydraulic fracturing (a.k.a., “fracking”) equipment such as Baker Hughes, Schlumberger (SLB), and Halliburton (HAL). When this attribute-specific shock occurs, you and the other value investors studying Schlumberger, Halliburton, and Baker Hughes start buying up shares. As an arbitrageur, when I see the resulting price increases in these 3 stocks, and I have to ask myself: “What should my next trade be?” When there have been fewer than \( N^{*} \) price changes in the petroleum industry, you have an informational monopoly. I can’t tell whether you are trading on a Houston, TX-specific shock or a fracking-specific shock since all 3 of these companies share both these attributes. I need to see at least \( N^{*} \) observations in order to recognize the pattern you’re trading on. The model in Section 3 reveals that you don’t need to alter your trading behavior in a large market before \( N^{*} \) transactions have occurred; however, once this threshold has been reached, I will recognize that Baker Hughes’ positive returns are due to a fracking shock and not a Houston, TX shock and rush into the market.

A common first response to this setup is: “The petroleum equipment industry does have
that many payout-relevant attributes. Can’t I just learn everything I need to know by tracking crude oil prices?” No. Figure 5 gives a sense of the number of different kinds of shocks that affect the petroleum equipment and services industry by plotting the frequency of Wall Street Journal article subjects. Just over the course of the last 3 years the industry has been hit with shocks relating to currency risk, liquid natural gas, fracking, regulatory changes, and merger activities. There are clearly many different attributes that people pay attention to when covering this industry—i.e., $Q \gg 1$. What’s more, very few of the $Q$ possible attributes matter each month. e.g., Figure 6 shows that only around 10% of all the descriptors in the Wall Street Journal articles about the petroleum equipment and services industry over the period from January 2011 to December 2013 are used each month. Thus, only a small fraction of the many possible attributes realize shocks each period—i.e., $K \ll Q$. To be sure, the price of crude oil is an important indicator for the industry. The term ‘Crude oil prices’ occurs in roughly half the articles; however, Figure 7 shows that this is the exception not the rule. Zipf’s law governs the distribution of petroleum industry article subjects in the Wall Street Journal. A general consequence of power law distributions in article topics as studied in Gabaix (2009) is that there are many equally likely and very obscure shocks.

How does the signal recovery bound apply in this situation? Suppose that only $K = 20$ attributes out of a possible $Q = 200$ attributes realized a shock in the previous period, and you discovered 1 of them through your research. How long does your informational monopoly last? Proposition 1 says that arbitrageurs need at least:

$$N^*(200, 20) \approx 20 \cdot \log^{(200/20)} = 46.05$$

(30)

price changes to identify which 20 of the 200 possible payout-relevant attributes in the petroleum industry realized a shock. If it takes you around 1 day to materially increase your position (i.e., there is 1 price change per day), then you have only 1 day to build
Zipf’s Law for Descriptive Tags in WSJ

**Figure 7.** Distribution of petroleum industry descriptive tags across Wall Street Journal articles. *x*-axis, log-scale: Fraction of all Wall Street Journal articles over the time period from January 2011 to December 2013 with a particular tag catalogued using ProQuest. *y*-axis, log scale: Probability that a randomly selected tag from Figure 5 occurs in a fraction *x* or larger of all articles. Reads: “Only 5% of the tags occur in more than 5% of all articles.”

up a position before the rest of the market catches on since there are 51 publicly traded companies in the petroleum equipment and services industry according to Yahoo! Finance. Baker Hughes, Schlumberger, and Halliburton will all realize positive returns on that first day due to value investor trading. The remaining 19 other shocks will also lead to very positive or negative returns for a small number of stocks on that first day, and by 4pm EST the variation in returns among the 51 stocks will be enough to pinpoint exactly which of 20 of the 200 possible attributes realized a shock prior to the start of trading. By contrast, if you had found a similarly rare shock in the major integrated oil and gas industry which has only 26 companies, then you would have nearly 2 days to trade before the rest of the market discovered the underlying fracking shock.

4.2. **Root Explanations.** This subsection gives a real world situation where it might not be worth it for market-wide arbitrageurs to wait long enough to identify the underlying attribute-specific shocks. e.g., it gives an example where the market is in the region of Figure 4 where $\eta \approx 0^+$. Musto, Nini, and Schwarz (2014) document “a large and systematic discrepancy among off-the-run Treasury securities: bond prices traded as much as 5% below otherwise identical notes.” Figure 8 shows that if on December 15th, 2008 you had gone long 30 year Treasury bonds maturing on February 15th, 2015 and short 10 year Treasury notes maturing at the same date and held this position for 1 month you would have earned a 15% annualized return. Suppose you’re a young bond trader sitting at your desk on the morning of December 15th, 2008 with $50\text{M}$ to invest. You know that if the spread continues to close, then you could make returns of 1.25% per month (i.e., 15% per year) by putting on a bond minus note trade:

$$r_{\text{BmN},t} = 1.25\% + \epsilon_t \quad \text{with} \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\epsilon)$$

(31)
where $\sigma^2 = 0.50\%$ per month to make the algebra neat.

Your big worry is that you have no idea why the spread opened up in the first place or what is causing it to close right now. i.e., you know that:

$$E[r_{BmN,t}] = \sum_{q=1}^{Q} \beta_q \cdot x_{BmN,q} = 1.25\%$$

(32)

but you don’t know which elements of $\beta$ are non-zero. As a result, you can only exploit this pricing error in the off-the-run Treasury market. It’s likely that whatever caused spreads to widen and then collapse in the Treasury market also caused other bond spreads to do the same thing—e.g., with corporate CDS and cash as in Bai and Collin-Dufresne (2011) or Treasury bonds and TIPS as in Fleckenstein, Longstaff, and Lustig (2013). Only trading the anomaly in the Treasury market exposes you to idiosyncratic risk of $\sigma^2/2$ per month that you would otherwise diversify away. e.g., looking at Figure 8 it’s clear that arbitrageurs tried to exploit this gap in August and October 2008 only to have spreads widen even further.

You have 2 options. On one hand, you could put on the trade using all of your available funds without understanding the underlying economics. This approach gives you an annualized Sharpe ratio of:

$$SR = \sqrt{12} \times \left( \frac{0.0125}{\sqrt{0.005}} \right) \approx 0.61$$

(33)

on your $50$ $\text{M}\$ initial investment. This isn’t a bad outcome, but it’s not substantially higher than the annualized Sharpe ratio on the market. On the other hand, you could try to use some of your $50$ $\text{M}\$ budget to identify the shocked attributes by placing test trades. e.g., place a $1$ $\text{M}\$ bet that will pay off if liquidity concerns are driving the spread, another $1$ $\text{M}\$ bet that will pay off if differing funding costs are driving the spread, another $1$ $\text{M}\$ bet that
will pay off if inventory risk for 30 Treasury bond holders is driving the spread, etc. This is analogous to the approach followed by the market-wide arbitrageurs in Section 3 with an opportunity cost of delaying $c = 1\text{M}$. 

The US Treasury market is a complex place subject to all sorts of different shocks. Suppose for simplicity that there are just $K = 2$ root explanations driving the spread, the market has $Q = 10^4$ payout relevant attributes, and it takes 1 day to see a trade pan out. In this world, it would take at least:

$$N^*(Q, K) = 2 \cdot \log(10^4/2) \approx 17$$

(34)

test trades at $1\text{M}$ a pop to identify which underlying attributes realized a shock. By experimenting for a day, you would earn a Sharpe ratio of:

$$\text{SR} = \sqrt{12} \times \left( \frac{0 \cdot 17/50 + 0.0125 \cdot (50 - 17)/50}{\sqrt{0.005 \cdot 17/50 + 0.0025 \cdot (50 - 17)/50}} \right) \approx 0.49$$

(35)

which is actually lower than if you had not tried to identify the fundamental shocks.

Thus, even though you could in principle back out the underlying cause of the matched maturity Treasury bond vs. note spread, it isn’t worth doing so. You are better off as a trader just placing your bet and letting the chips fall where they may. Consequently, the same sort of pricing error may repeatedly arise in different guises because traders never have the proper incentives to investigate the root cause. What’s more, there seems to be at least anecdotal evidence of this phenomenon in real world markets. e.g., the co-CEO of Renaissance Technologies, Robert Mercer, has pointed out that “some signals that make no intuitive sense do indeed work... The signals that we have been trading without interruption for 15 years make no sense, otherwise someone else would have found them.”

4.3. **House-Specific Price Index.** This subsection asks the question: “Is is possible to create a meaningful price index for a single house?” e.g., Zillow research computes a house price index at aggregation levels ranging all the way up to state-wide and all the way down to neighborhood-specific. Is there some smallest level of aggregation? Or, would a house-specific index be just a really noisy signal about the true underlying house price? The existence of the signal recovery bound described in Section 2 suggests that is no information content in a single-house index. Such an index wouldn’t just be noisy; it would be meaningless.

To see why, think about how you would create a standard city-level hedonic house price index for, say, Berkeley, CA. First, you must collect price and house attribute data from the past $N$ sales. e.g., the 3rd recent sale in Berkeley, CA might have taken place for a price

---


of $1.2\text{M\$}, and the house might have had granite countertops and a half-circle driveway but lacked a pool. After collecting the data, you would then estimate how people in Berkeley, CA have changed their preferences for houses with each of the $Q$ different attributes. i.e., you would estimate the system:

$$\Delta p_n = p_n - E[p_n] = \sum_{q=1}^{Q} \beta_q \cdot x_{n,q} + \epsilon_n \quad \text{where} \quad K = \|\beta\|_{\ell^0} = \sum_{q=1}^{Q} 1_{\{\beta_q \neq 0\}} \quad (36)$$

where the answers to the yes/no questions: “Does this house have granite countertops?”, “Does this house have a half-circle driveway?”, “Does this house have a pool?”, etc. . . represent each house’s exposure, $x_{n,q}$, to each of the $Q$ possible attributes and $\epsilon \sim \text{iid} \sim N(0, \sigma^2)$ denotes purely idiosyncratic fluctuations in particular sale prices. In practice, home buyers only change their preferences for $K \ll Q$ of the many possible attributes each month, so Equation (36) is identical to the ones studied above.

When looking at entire cities, it seems very likely that the number of sales each month exceeds the signal recovery bound, $N^*(Q,K)$, required to estimate which of the elements in $\beta$ have changed. e.g., because there are enough sales at the city-level to reach the signal recovery bound, it is feasible to look at past prices and say that: “Home buyers in Berkeley now value having a half-circle driveway more than they used to; whereas, home buyers in San Francisco still value having a half-circle driveway the same way but value living in a good school district way more than they used to.”

However, at some level of granularity there won’t be enough sales to identify which elements in $\beta$ are causing house prices to fluctuate. At some level of granularity, there won’t be enough sales to reach the signal recovery bound. e.g., Figure 9 shows a map of recent home sales in Berkeley, CA with the dashed lines outlining a pair of adjacent neighborhoods. There was 1 recent sale in the neighborhood of North Berkeley and 2 recent sales in the neighborhood of Central Berkeley. Looking at Central Berkeley, 2 sales is fewer than the signal recovery bound unless each house only $Q = 3$ attributes! Looking at North Berkeley takes this point to the extreme with only 1 sale. There isn’t enough information to say: “Home buyers in North Berkeley now value have a half-circle driveway more than they used to; whereas, home buyers in Central Berkeley now value living in the better school district more.” It is possible to compute whether prices have been moving up or down via a repeat sales index as in Case and Shiller (1987) and Case and Shiller (1989), but you can’t tell why. To be sure, everyone in Berkeley, CA will have their own beliefs about what the price of any particular house in North Berkeley or Central Berkeley should be; however, if any 2 people’s beliefs don’t agree, there just isn’t enough data to arbitrate their disagreement.

This insight has broader implications for financial econometrics. Many important asset
markets involve infrequently traded assets. e.g., think about private equity stakes (see Korteweg and Sorensen (2013)), thinly traded stocks (see Dimson (1979) and Easley, Kiefer, O’Hara, and Paperman (1996)), and OTC bond markets (see Harris and Piwowar (2006)). Existing analysis has assumed that rarely traded assets have some true price which evolves over time, and each sale is a sample from this fundamental price process. i.e., existing work assumes there is an asset-specific price index. It is no longer obvious that such a fundamental price process exists.

5. Conclusion

This paper asks a simple question, “How many price changes do traders need to see in order to tease out which financial shocks have occurred?”, and derives a surprising answer via tools from the compressed sensing literature. If fewer than $N^*$ transactions have occurred, then knowledge about which shocks took place is inherently local since price changes aren’t sufficient to broadcast this information. Traders must use some other information source in addition to price changes—e.g., word of mouth. After characterizing this signal recovery
bound, $N^*$, I show how it can affect equilibrium asset prices by couching it in a Kyle (1985)-
type model. Finally, I give evidence of the broad applicability of this bound by providing
additional examples of how it impacts traders in several different asset markets. While the list
of examples in this paper is suggestive, it is by no means comprehensive. It has not escaped
the author’s notice that the signal recovery bound has implications for understanding the
prevalence of seemingly redundant securities, the existence of pay-to-trade market structures,
and the interpretation of event studies.

The goal of the paper is not to suggest that the signal recovery bound is the only friction
traders face. It is not. Nevertheless, the finite information capacity of a finite number of
price changes does appear to be a binding constraint in many important market settings, and
it is a constraint that is not captured in the standard noisy rational expectations setup of
Grossman and Stiglitz (1980) and Kyle (1985). In these models traders are only interested in
finding out the expected value of a single security; by contrast, in the current paper traders
are interested in discovering the attribute-specific shocks underlying the changes in asset
values. A key implication of this new setting is that there is a link between the breadth of
the market (i.e., the number of payout-relevant attributes) and the profitability of hunting
for arbitrage opportunities (i.e., the accuracy of prices). This connection presents promising
avenues for future research.
References


Appendix A. Proofs

**Proof** (Proposition 1, Sufficiency). Assume that $K, N, Q \to \infty$, $K/q \to 0$, $(N - K) \cdot \beta \to \infty$, and $\beta = 1/\sqrt{\pi}$. I show that if for some constant $a > 0$:

$$N > a \cdot K \cdot \log(Q/K)$$

then there exists an asymptotically reliable selection rule. The proof of sufficiency proceeds in 4 steps. First, I lay out what it would mean for an optimal selection rule to choose the $K$ wrong attributes in the form of an inequality. Then, I make use of the normality of both $x_{n,q}$ and $\epsilon_n$ to reinterpret this bound analytically. Next, I use the assumptions about the growth rates of $N, K,$ and $Q$ to evaluate this inequality. Finally, I sum over all choices of errors to generate the desired result.

1. **Reconstruction error.** I start by defining what it means for a selection rule to make a mistake. Let $J \subset Q$ denote a subset of $Q$ with $J$ elements, and define $f(\cdot)$ as the reconstruction error of this set:

$$f(J; \Delta p, X) = \|\Delta p - X_J \beta_J\|_2^2$$

e.g., for the true subset of shocked attributes, $K$, the reconstruction error is:

$$\langle f(K; \Delta p, X) \rangle = \sigma^2$$

where the average is taken over all choices of $X$ and $\epsilon$. Suppose that you know exactly how many attributes have realized a shock, $K$; you just don’t know which $K$. In this setting, the optimal decoder, $\phi^*$, is given by:

$$\hat{K} = \phi^*(\Delta p, X) = \arg \min_{|J|=K} f(J; \Delta p, X)$$

Define the Diff[] operator as the difference between the reconstruction error using the set $J$ and the reconstruction error using the true subset $K$:

$$\text{Diff}[J] = f(J; \Delta p, X) - f(K; \Delta p, X)$$

The optimal decoder chooses $J$ over $K$ if and only if:

$$0 > \text{Diff}[J] = \|\Pi_J^\perp (X_{K\setminus J} \beta_{K\setminus J} + \epsilon)\|_2^2 - \|\Pi_K^\perp \epsilon\|_2^2$$

where the matrices $\Pi_J$ and $\Pi_J^\perp$ are defined as:

$$\Pi_J = X_J (X_J^\top X_J)^{-1} X_J^\top \quad \text{and} \quad \Pi_J^\perp = I - \Pi_J$$

i.e., if for the particular choice of attribute exposures, $X$, and random noise, $\epsilon$, there exists some subset of shocks $J$ with $|J| = K$ elements that happens to have a lower reconstruction error than the true subset, $f(J) > f(K)$:

$$\Pr[\hat{K} \neq K] = \Pr(\exists J \subset Q \text{ s.t. } |J| = K, \text{Diff}[J] < 0) \geq \Pr(\text{Diff}[J] < 0)$$

2. **Distributional assumptions.** Let’s now dig into this reconstruction error more deeply by taking advantage of the distributional assumptions. Let $H$ denote the number of attributes which the optimal selection rule misclassifies. i.e., for any $J = \phi^*(\Delta p, X)$:

$$|K \setminus J| = H$$

Using the fact that $x_{n,q} \overset{iid}{\sim} N(0, 1)$ and $\epsilon_n \overset{iid}{\sim} N(0, \sigma^2)$, it is possible to show that for a
fixed $H$ with $1 \leq H \leq K$:
\[
\Pr(\text{Diff}[\mathcal{J}] < 0) \leq \exp \left\{ -\frac{(N-K) \cdot \|eta_{\mathcal{K}\setminus\mathcal{J}}\|_2^2}{12 \cdot (\|eta_{\mathcal{K}\setminus\mathcal{J}}\|_2^2 + 4)} \right\} \\
+ \exp \left\{ -\frac{H}{4} \cdot \left( \left( \frac{N-K}{4} \right) \cdot \frac{\|eta_{\mathcal{K}\setminus\mathcal{J}}\|_2^2}{H} - 1 \right)^2 \right\}
\]

Clearly, we have that:
\[
\|eta_{\mathcal{K}\setminus\mathcal{J}}\|_2^2 = H \cdot \beta^2
\]
since $\beta_q = \beta$ for all $q \in \mathcal{K}$. Thus, I can simplify the bound above to:
\[
\Pr(\text{Diff}[\mathcal{J}] < 0) \leq \exp \left\{ -\frac{(N-K) \cdot H \cdot \beta^2}{12 \cdot (H \cdot \beta^2 + 8)} \right\}
\]

(3) Asymptotic evaluation. Next, I exploit the asymptotic assumption of $(N-K) \cdot \beta_{\text{min}}^2 \to \infty$ to further simplify this inequality. Specifically, both the terms in the equation above are upper bounded by $\exp \left\{ -\frac{(N-K) \cdot H \cdot \beta^2}{12 \cdot (H \cdot \beta^2 + 8)} \right\}$. The first term obviously satisfies this bound. For the second term, you can show that:
\[
-\frac{(N-K) \cdot H \cdot \beta^2}{12 \cdot (H \cdot \beta^2 + 8)} \geq -\frac{H}{12} \cdot \left( \left( \frac{N-K}{4} \right) \cdot \beta^2 - 1 \right)^2
\]

Thus, we have a much simplified expression:
\[
\Pr(\text{Diff}[\mathcal{J}] < 0) \leq \exp \left\{ -\frac{(N-K) \cdot H \cdot \beta^2}{12 \cdot (H \cdot \beta^2 + 8)} \right\}
\]

(4) Summing over possibilities. Finally, I sum over all possible errors that the optimal selection rule could possibly make in order to get:
\[
\text{Err}[\phi] \leq \sum_{h=1}^{K} \binom{K}{h} \cdot \binom{Q-K}{h} \cdot \exp \left\{ -\frac{(N-K) \cdot h \cdot \beta^2}{12 \cdot (h \cdot \beta^2 + 8)} \right\}
\]
since the number of misclassified attributes could be any value $1 \leq H \leq K$. The binomial coefficients are nasty to work with directly, so I use the standard bounds:
\[
\log \binom{K}{k} \leq k \cdot \log \left( \frac{K \cdot e}{k} \right) \quad \text{and} \quad \log \binom{Q-K}{k} \leq k \cdot \log \left( \frac{(Q-K) \cdot e}{k} \right)
\]

Using these inequalities, the log of the $h$th term is bounded above by:
\[
\log \left[ \binom{K}{h} \cdot \binom{Q-K}{h} \cdot e^{\frac{(N-K) \cdot h \cdot \beta^2}{12 \cdot (h \cdot \beta^2 + 8)}} \right] \\
\leq h \cdot (2 + \log(k/h) + \log((Q-K)/h)) - \frac{(N-K) \cdot h \cdot \beta^2}{12 \cdot (h \cdot \beta^2 + 8)}
\]
If we want $\text{Err}[\phi] \to 0$, then we want each of these logarithmic terms to be as negative as possible. Requiring the $h$th term to be negative means that:

$$N \geq K + 12 \cdot (h + 8/\beta^2) \cdot (2 + \log(K/h) + \log [(Q-K)/h])$$

Since $K/Q \to 0$ and $\beta = 1/\sqrt{\kappa}$, the sum over all possible right hand side values is dominated by the $h = K$ term so that if for some $a > 0$:

$$N > a \cdot K \cdot \log((Q-K)/K) \approx a \cdot K \cdot \log(Q/K)$$

there will be no negative terms which is the desired result.

\[\square\]

**Proof** (Proposition 1, Necessity). Assume that $K, N, Q \to \infty$, $K/Q \to 0$, $(N-K) \cdot \beta \to \infty$, and $\beta = 1/\sqrt{\kappa}$. I show that if for some constant $a' \in (0, a]$:

$$N < a' \cdot K \cdot \log(Q/K)$$

then there does not exists an asymptotically reliable selection rule. The proof proceeds in 2 steps. First, I use Fano’s inequality to give a lower bound on the probability of making an error in high-dimensional hypothesis testing situation. See Cover and Thomas (1991) for an overview of Fano’s inequality. Then, I make use of the normality of both $x_{n,q}$ and $\epsilon_n$ to give an tail bound on the probability of not making any errors.

(1) Fano’s inequality. I begin by giving a lower bound on the probability of selecting a the wrong subset of attributes via Fano’s inequality. Let $S = \binom{Q}{K}$ denote the number of attribute subsets $\mathcal{J} \subset \mathcal{Q}$ of size $|\mathcal{J}| = K$ and index each of these subsets with $\mathcal{J}_s$ for $s = 1, 2, \ldots, S$. Suppose that $P_s$ denotes a multivariate normal distribution with:

$$P_s = N(\beta \cdot X_{\mathcal{J}_s} \mathbf{1}, \mathbf{I}) \quad \text{for } s = 1, 2, \ldots, S$$

where $\mathbf{1}$ denotes a $(K \times 1)$-dimensional vector of 1s, and $\mathbf{I}$ denotes the $(K \times K)$-dimensional identity matrix. The optimal selection rule searches over all $S$ attribute subsets and tries to solve the program:

$$\min_{s=1,2,\ldots,S} \|y - \beta \cdot X_{\mathcal{J}_s} \mathbf{1}\|_{l_2}^2 = \min_{s=1,2,\ldots,S} \|\beta \cdot (X_K - X_{\mathcal{J}_s}) \mathbf{1} + \epsilon\|_{l_2}^2$$

Fano’s inequality states (after slight modification) that for a family of $S$ distributions $\{P_1, P_2, \ldots, P_S\}$:

$$\text{Err}[\phi] = \Pr[\tilde{K} \neq K] \geq 1 - \left(\frac{1}{2\pi} \cdot \sum_{s,s'=1}^S \text{Div}(P_s|P_s') + \log 2}{\log(S-1)}\right)$$

where $\text{Div}(\cdot|\cdot)$ denotes the Kullback-Leibler divergence. Plugging in the form of the optimization problem to characterize the Kullback-Leibler divergence then gives:

$$\text{Err}[\phi] \geq 1 - \frac{1}{2} \left(\frac{1}{2\pi} \cdot \sum_{s,s'=1}^S \|\beta \cdot (X_{\mathcal{J}_s} - X_{\mathcal{J}_s'}) \mathbf{1}\|_{l_2}^2 + 2 \cdot \log 2}{\log(S-1)}\right)$$

The remainder of the proof involves massaging the right hand side of the equation above to keep $\text{Err}[\phi] > 0$—i.e., the necessary condition for asymptotic reliability.
(2) Distributional assumptions. In order for $\text{Err}[\phi] > 0$, it has to be the case that:

$$1 > \frac{1}{2 \cdot S^2} \cdot \sum_{s,s'=1}^{S} \| \beta \cdot (X_{J_s} - X_{J_{s'}}) \|_{\ell_2}^2.$$ 

For any given pair of subsets $(J_s, J_{s'})$ define the random variable:

$$Z_{s,s'} = \| \beta \cdot (X_J - X_{J_{s'}}) \|_{\ell_2}^2$$

Using the normality of each of the attribute exposures, $x_{n,q} \sim N(0, 1)$, it’s possible to show that $Z_{s,s'}$ adheres to a scaled $\chi^2_N$ distribution:

$$Z_{s,s'} \sim 2 \cdot \beta^2 \cdot (K - |J_s \cap J_{s'}|) \cdot \chi^2_N$$

where $|J_s \cap J_{s'}|$ denotes the size of the set difference between the subsets $J_s$ and $J_{s'}$. e.g., if there are $K = 4$ shocked attributes and $J_s = \{1, 2, 5, 9\}$ while $J_{s'} = \{1, 3, 5, 9\}$, then $|J_s \cap J_{s'}| = 1$. Using the tail bound for a $\chi^2_N$ distribution, we see that:

$$\text{Pr} \left[ \frac{1}{S^2} \cdot \sum_{s \neq s'} Z_{s,s'} \geq 4 \cdot \beta^2 \cdot K \cdot N \right] \leq 1/2$$

Thus, at least half of the $S$ different distributions $\{P_1, P_2, \ldots, P_S\}$ satisfy the bound:

$$\frac{1}{2 \cdot S^2} \cdot \sum_{s,s'=1}^{S} \| \beta \cdot (X_{J_s} - X_{J_{s'}}) \|_{\ell_2}^2 \leq \frac{4 \cdot \beta^2 \cdot K \cdot N}{\log(S - 1)}.$$

Thus, the distributional assumptions on the attribute exposures, $x_{n,q}$, imply that as long as $1 > (4 \cdot \beta^2 \cdot K \cdot N)/\log(S - 1)$, the error rate will remain bounded away from 0 implying that:

$$N > \frac{\log(S - 1)}{4 \cdot \beta^2 \cdot K}$$

is necessary for asymptotically reliable recovery. To make the formula match, note that:

$$\log(S - 1) \geq \frac{1}{2} \cdot \log N \geq \frac{K}{2} \cdot \log(Q/\kappa)$$

\[\square\]

**Proof** (Proposition 2). Value investors in asset $n$ solve the problem:

$$\pi^V_n = \max_{y^V_n} \mathbb{E} \left[ (v_n - p_n) \cdot y^V_n \mid v_n \right]$$

$$= \max_{y^V_n} \mathbb{E} \left[ (v_n - \lambda_n \cdot \{y^V_n + y^A_{n} + \epsilon_n\}) \cdot y^V_n \mid v_n \right]$$

Taking the first order condition with respect to $y^V_n$ yields:

$$0 = \mathbb{E} \left[ v_n - 2 \cdot \lambda_n \cdot y^V_n - \lambda_n \cdot y^A_{n} - \lambda_n \cdot \epsilon_n \mid v_n \right]$$

Evaluating the expectations operator gives:

$$0 = v_n - 2 \cdot \lambda_n \cdot y^V_n - \lambda_n \cdot y^A_{n}$$
so that:

\[ y_{VI}^n = v_n \times \begin{cases} 
\frac{1 - \lambda_* \theta_{Arb}}{2 \lambda_*} & \text{if } n > N^* \\
\frac{1}{2 \lambda_*} & \text{else}
\end{cases} \]

Market-wide arbitrageurs trading asset \( n > N^* \) solve the problem:

\[
\pi_{Arb}^n = \max_{y_{Arb}^n} E \left[ (v_n - p_n) \cdot y_{Arb}^n \middle| v_n \right] \\
= \max_{y_{Arb}^n} E \left[ (v_n - \lambda_* \cdot \left\{ y_{VI}^n + y_{Arb}^n + \epsilon_n \right\} \cdot y_{Arb}^n \middle| v_n \right]
\]

Taking the first order condition with respect to \( y_{Arb}^n \) yields:

\[
0 = E \left[ v_n - \lambda_* \cdot \left( y_{VI}^n + y_{Arb}^n - \lambda_* \cdot \epsilon_n \right) \middle| v_n \right]
\]

Evaluating the expectations operator gives:

\[
0 = v_n - \lambda_* \cdot y_{VI}^n - 2 \cdot \lambda_* \cdot y_{Arb}^n - \lambda_* \cdot \epsilon_n
\]

so that:

\[
y_{Arb}^n = \left( \frac{1 - \lambda_* \cdot y_{VI}^n}{2 \cdot \lambda_*} \right) \cdot v_n
\]

Clearly, for \( n \leq N^* \) we have the standard static Kyle (1985) solution:

\[
\lambda_0 = \frac{1}{(2 \cdot \sigma_e)} \quad \text{and} \quad \theta_{VI}^0 = \sigma_e
\]

with \( \text{Var}[v_n] = 1 \). When \( n > N^* \) things get more interesting. The market maker’s boundary condition implies that:

\[
p_n = \frac{\text{Cov}[v_n, y_n]}{\text{Var}[y_n]} \cdot y_n
\]

Thus, we have that:

\[
\lambda_* = \frac{\text{Cov}[v_n, y_{VI}^n + y_{Arb}^n + \epsilon_n]}{\text{Var}[y_{VI}^n + y_{Arb}^n + \epsilon_n]} = \frac{\text{Cov}[v_n, \theta_{VI}^* \cdot v_n + \theta_{Arb}^* \cdot v_n + \epsilon_n]}{\text{Var}[\theta_{VI}^* \cdot v_n + \theta_{Arb}^* \cdot v_n + \epsilon_n]} = \frac{\theta_{VI}^* + \theta_{Arb}^*}{(\theta_{VI}^* + \theta_{Arb}^*)^2 + \sigma_e^2}
\]

However, in equilibrium we know that \( \theta_{Arb}^* = \theta_{VI}^* \) since both agents have the exact same information leaving the system of 2 equations and 2 unknowns:

\[
\theta_* = \frac{1 - \lambda_* \cdot \theta_*}{2 \cdot \lambda_*} \quad \text{and} \quad \lambda_* = \frac{2 \cdot \theta_*}{4 \cdot \theta_*^2 + \sigma_e^2}
\]

Solving the system leads to the solution:

\[
\lambda_* = \sqrt{2}/(3 \cdot \sigma_e) \quad \text{and} \quad \theta_* = \frac{\sigma_e}{\sqrt{2}}
\]

\[\square\]

**Proof** (Proposition 3). Substituting in the equilibrium values into the value investor’s utility
function gives:

\[
E[\pi_n^{VI}] = \frac{N^*}{N} \cdot E_n^{VI} \left[ (v_n - \lambda_o \cdot \{\theta_o \cdot v_n + \epsilon_n\}) \cdot \theta_o \cdot v_n \mid n \leq N^* \right]
\]

\[
+ \frac{N - N^*}{N} \cdot E_n^{VI} \left[ (v_n - \lambda_\star \cdot \{2 \cdot \theta_\star \cdot v_n + \epsilon_n\}) \cdot \theta_\star \cdot v_n \mid n > N^* \right]
\]

\[
= \frac{\sigma_\epsilon}{\sqrt{2}} \cdot \left\{ \frac{1}{\sqrt{2}} \cdot \frac{N^*}{N} + \frac{1}{3} \cdot \left( 1 - \frac{N^*}{N} \right) \right\}
\]

Substituting in the equilibrium values into the arbitrageur’s utility function gives:

\[
E[\pi^{Arb}] = (N - N^*) \cdot \frac{\sigma_\epsilon}{3 \cdot \sqrt{2}} - c \cdot N^*
\]

Evaluating the limit as \( N, Q, K \to \infty, \frac{K}{Q} \to 0, \) and \( \frac{\Delta N}{N^*} \to \eta \) gives:

\[
E[\pi_n^{VI}] = \sigma_\epsilon \cdot \left( \frac{3 + \sqrt{2} \cdot \eta}{6 \cdot (1 + \eta)} \right)
\]

and:

\[
\pi^{Arb} = \left( \frac{\eta \cdot \sigma_\epsilon - 3 \cdot \sqrt{2} \cdot c}{3 \cdot \sqrt{2}} \right) \cdot N^*
\]

Equating each agent type’s expected utility yields a solution for the equilibrium learning style choice:

\[
E[\pi_n^{VI}] \cdot \frac{1}{\omega/N} = E[\pi^{Arb}] \cdot \frac{1}{1 - \omega} \quad \text{with} \quad \omega = \frac{\sigma_\epsilon}{3} \cdot \frac{3 \cdot \sqrt{2} + 2 \cdot \eta}{\sqrt{2} \cdot (\sigma_\epsilon - c) + \sigma_\epsilon \cdot \eta}
\]