News or Noise? The Missing Link*

Ryan Chahrour
Boston College

Kyle Jurado
Duke University

May 17, 2017

Abstract

The literature on belief-driven business cycles treats news and noise as distinct representations of people's beliefs. We prove they are empirically the same. Our result lets us isolate the importance of beliefs as an independent source of fluctuations. Using three prominent estimated models, we show that existing research understates this importance. Our result implies that structural vector autoregression analysis can be applied to models with either news or noise. We demonstrate this in U.S. data, and find that productivity accounts for 14% of consumption fluctuations, of which only a small portion is due to future shocks; the rest is noise.

JEL classification: D84, E32, C31
Keywords: News, noise, business cycles, structural vector autoregression

*Chahrour: Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 (e-mail: ryan.chahrour@bc.edu); Jurado: Department of Economics, Duke University, 419 Chapel Drive, Durham, NC 27708 (e-mail: kyle.jurado@duke.edu).

1 Introduction

A large literature in macroeconomics has argued that changes in people's beliefs about the future can be an important cause of economic fluctuations.[1] This idea, which dates at least to Pigou (1927), has been formalized in two ways. In the first way, which we call a "news representation," people perfectly observe part of an exogenous fundamental in advance. By way of analogy, this is like learning today that in next week's big game your favorite team will certainly win the first half. You don't know whether they will win the game, which is ultimately what you care about, because you are still unsure how the second half will turn out. In the second way, which we call a "noise representation," people imperfectly observe an exogenous fundamental in advance. This is like your friend telling you that he thinks your team will win next week's game. He follows the sport more than you do, and is often right, but sometimes he gets it wrong.

Much of the literature emphasizes the differences between these two ways of representing people's beliefs.[2] For example, in models with news, people have full information and shocks are perfectly anticipated; in models with noise, people have imperfect information and shocks are not perfectly anticipated. It has been suggested that models with noise shocks may be more theoretically flexible, and require weaker assumptions regarding the timing of information arrival. Others argue that models with news shocks may be easier to estimate using semi-structural empirical methods, which rely on fewer theoretical assumptions. Some studies include both news and noise shocks in the same model and attempt to determine which is more important.

In this paper, we argue that news and noise representations are more closely linked than the literature has recognized. Specifically, we prove that these two information structures are observationally equivalent. This means that even given an ideal data set with complete observations of exogenous fundamentals *and* people's beliefs about those fundamentals, it would be impossible to tell them apart. It therefore follows that neither representation requires stronger modeling assumptions for theoretical work, or greater reliance on a model's structural details for empirical work.

Our main result is a representation theorem, which says that fundamentals and people's beliefs about them always have both a news representation and a noise representation. This implies that associated with every noise representation is an observationally equivalent news representation and vice versa. We present a constructive proof of the theorem using Hilbert

[1] Throughout the paper, we use the words "beliefs," "expectations," and "forecasts" as synonyms.

[2] This emphasis is often implicit in discussions of news and noise. Relatively explicit examples include Sections 2 and 4.2.3 of Beaudry and Portier (2014), Sections 5 and 6 of Lorenzoni (2011), Sections II.B and II.C of Blanchard et al. (2013), and the introduction of Barsky and Sims (2012).

space methods. Because it is constructive, our proof also provides a method for explicitly deriving the mapping from one representation to another. We compute this mapping in closed form for several models of interest from the literature.[3]

The main step in moving from noise to news amounts to finding the Wold representation of the noise model. This is because the shocks in the news representation are static rotations of the Wold innovations implied by the noise representation. Because the Wold innovations are contained in the space spanned by the history of variables that people observe, the news representation is a way of writing models with noise "as if" people have perfect information.[4] To move in the opposite direction, from news to noise, the idea is to reverse engineer the signal extraction problem that generates a given Wold representation. The challenge is to ensure that the noise shocks in that signal extraction problem are independent of fundamentals at all leads and lags, and that they capture all the non-fundamental variation in beliefs.

Beyond clarifying the link between news and noise, our representation theorem sheds new light on two important questions in the literature. The first is: how important are beliefs as an independent source of fluctuations? The second is: can structural vector autoregressions (VAR) analysis be applied to models with noise shocks?

The first of these two questions is central to the literature on belief-driven business cycles. However, existing studies that either use models with only news shocks or some combination of news and noise shocks cannot answer it. The reason is that news shocks mix the fluctuations due to beliefs with the contribution due to fundamentals. News shocks can change beliefs on impact without any change in current fundamentals, but they are tied by construction to changes in future fundamentals. Beliefs change today, and on average fundamentals change tomorrow. But which is more important, the change in beliefs or the subsequent change in fundamentals?

To isolate the independent contribution of beliefs, it is necessary to disentangle the effects due purely to expected changes in fundamentals from the consequences of their actual realizations.[5] One way to do this is to first find a noise representation of the news-shock model, and then consider the importance of noise shocks. Noise shocks isolate precisely those movements in beliefs that are independent of fundamentals at all horizons. Our representation theorem ensures that it is always possible to do this, and our constructive proof provides a procedure for doing so.

[3]For a discussion of representation theory in the context of covariance-stationary stochastic processes, including a statement of the famous theorem of Wold (1938), see Sargent (1987), ch. XI.

[4]A related result is Lemma 2 of Blanchard et al. (2013), which shows that their information structure has an observationally equivalent full information representation with correlated shocks.

[5]This point has been emphasized in the literature. For example, see the discussion in Section IV.A of Barsky, Basu, and Lee (2015), as well as the recent paper by Sims (2016).

We use our result to compute the independent contribution of beliefs implied by three different quantitative models of U.S. business cycles. The three models come from Schmitt-Grohé and Uribe (2012), Barsky and Sims (2012), and Blanchard et al. (2013). These models all appear to have very different information structures, which — combined with differences in the rest of the physical environment, estimation procedure, and data sample — has made it difficult to compare results across models. By allowing us to isolate the independent contribution of beliefs in each model, our representation theorem provides a way of coherently comparing them. We use the exact models and estimated parameters from the original papers. Because news and noise representations are observationally equivalent, the likelihood functions are the same under either representation.

In all three cases, the importance of independent fluctuations in beliefs has been understated. In the model of Schmitt-Grohé and Uribe (2012), there is no shock labeled "noise," but the implicit contribution of noise shocks is between 3% and 11% depending on the variable. In the model of Barsky and Sims (2012), noise shocks are responsible for 9% of the fluctuations in consumption, which is almost an order of magnitude larger than the original estimate of 1%. In the model of Blanchard et al. (2013), the contribution of noise to consumption is 57%, compared to the originally reported value of 44%.[6]

We also use the noise representation to isolate the contribution of future fundamentals relative to current and past fundamentals. Consistently across all three models, we find that future fundamentals are much less important than current and past fundamentals. For example, in the model of Barsky and Sims (2012), future fundamentals are responsible for less than 0.5% of consumption fluctuations, while current and past fundamentals are responsible for over 80%. Future fundamentals matter the most in the model of Blanchard et al. (2013), but even then they drive less than 7% of consumption fluctuations.

Interestingly, we find that future fundamentals are not very important despite the fact that all three models assign a fairly large role to news shocks. The reason is that, in addition to mixing fluctuations due to beliefs and fundamentals, news shocks *also* mix fluctuations due to past, present, and future fundamentals. Current news shocks reflect changes in future fundamentals, but past news shocks show up as changes in current fundamentals. If a model is not sufficiently "forward-looking," it may be that news shocks matter mainly through this second channel. It turns out that this is the case for all three models we consider.

The second question our representation theorem helps to answer is whether structural VAR analysis can be applied to models with noise shocks. A common view is that they

[6] Clearly, there remains substantial disagreement across models regarding the overall extent to which beliefs are important. We do not go into the reasons for that disagreement in this paper, but Section III.B of Barsky and Sims (2012) has a good discussion.

cannot.[7] However, we show that structural VARs can be applied to models with either news or noise shocks. Given any reduced-form VAR representation of fundamentals and beliefs, we can always compute its news or noise representation. The key step in our argument is to explain why "invertibility" is not a necessary condition for using VARs. True, the noise shocks will always depend on future reduced-form VAR shocks, but that is not a problem. In any sample of data, time is symmetric; we can look backward into the past from one end or forward into the future from the other.

To illustrate how structural VAR analysis can be used to think about noise shocks, we perform two exercises. First, we simulate an analytically convenient theoretical model of consumption determination with noise shocks, and show how structural VAR analysis can accurately uncover the effects and importance of both noise and fundamental shocks. Second, we apply the same procedure to a sample of postwar U.S. data on consumption and productivity. We find that less than 15% of the business-cycle variation in consumption can be attributed to productivity shocks, with all remaining fluctuations attributed to noise shocks. This finding represents a challenge for theories of consumption determination that rely primarily on beliefs about productivity. According to such theories, beliefs about productivity must be fluctuating in ways that are mostly unrelated to productivity itself.

The literature on both news and noise shocks is large. In the noise literature, Lorenzoni (2009), Angeletos and La'O (2013), and Benhabib et al. (2015) have explored models in which dispersed information across agents can generate fluctuations in beliefs that are independent of aggregate fundamentals; we restrict our analysis to cases with a single, representative information set. In the news literature, Cochrane (1994), Beaudry and Portier (2006), and Beaudry and Lucke (2010) all provide VAR-based evidence pointing to an important role for news, and some empirical DSGE studies not cited above, including Forni et al. (2014) and Christiano et al. (2014), have estimated large roles for such shocks. Leeper and Walker (2009) and Leeper et al. (2013) explore how the specification of news processes alters the effects of news shocks on the dynamics of endogenous variables. Other related papers include Jaimovich and Rebelo (2009), Beaudry et al. (2011), Lorenzoni (2011), Barsky and Sims (2011), Born et al. (2013), Kurmann and Otrok (2013), and Jinnai (2014).

2 Observational Equivalence

News and noise representations are two different ways of describing economic fundamentals and people's beliefs about them. "Fundamentals" are stochastic processes capturing exogenous changes in technology, preferences, endowments, or government policy. Throughout

[7] This is the main methodological point of Blanchard et al. (2013).

this section, fundamentals are summarized by a single scalar process $\{x_t\}$. People's decisions depend on expected future realizations of x_t, so both representations specify what people can observe at each date and how they use their observations to form beliefs about the future.

The main result of the paper, which is presented in this section, is a representation theorem linking news and noise representations. The first subsection presents the result in a simple example with news or noise regarding fundamentals only one period in the future while the second subsection presents the more general result.

2.1 Simple Example

In the simplest of news representations, x_t is equal to the sum of two shocks, $a_{0,t}$ and $a_{1,t-1}$, which are independent and identically distributed (i.i.d.) over time, and which are independent of one another:

$$x_t = a_{0,t} + a_{1,t-1}, \qquad \begin{bmatrix} a_{0,t} \\ a_{1,t} \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_{a,0}^2 & 0 \\ 0 & \sigma_{a,1}^2 \end{bmatrix}\right). \tag{1}$$

At each date t, people observe the whole history of the two shocks up through that date, $\{a_{0,\tau}, a_{1,\tau}\}$ for all integers $\tau \leq t$. Their beliefs regarding fundamentals are rational; the probabilities they assign to future outcomes are exactly those implied by system (1). The shock $a_{1,t}$ is a news or anticipated shock because people see it at date t but it doesn't affect the fundamental until date $t+1$. The shock $a_{0,t}$ is a surprise or unanticipated shock.

Now consider instead a noise representation. The fundamental variable x_t is i.i.d. over time, and there is a noisy signal of the fundamental one period into the future:

$$s_t = x_{t+1} + v_t, \qquad \begin{bmatrix} x_t \\ v_t \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}\right). \tag{2}$$

At each date t, people observe the whole history of fundamentals and signals up through that date, $\{x_\tau, s_\tau\}$ for all integers $\tau \leq t$. Even though people only have imperfect information about x_{t+1}, their beliefs are nevertheless still rational. The shock v_t is a noise or error shock because it affects beliefs but is totally independent of fundamentals.

Our point is that these two representations are observationally equivalent. But before making that point, it is important to be clear about what types of things we are considering to be "observable." To be concrete, imagine an econometrician who is able to observe the entire past, present, and future history of the fundamental process $\{x_t\}$, along with the entire past, present, and future history of people's subjective beliefs regarding $\{x_t\}$. More concisely, we will say that the econometrician observes "fundamentals and beliefs." All of our results are stated from the perspective of such an econometrician, and are to be understood with respect to those observables.

An important feature of our concept of equivalence is that we treat beliefs, as well as fundamentals, as observable. We take this approach for three reasons. First, it is a stronger condition; observational equivalence with respect to a larger set of observables implies observational equivalence with respect to any smaller set of those observables. Second, beliefs are observable in economics, in principle. Beliefs may be measured directly, using surveys, or indirectly, using the mapping between beliefs and actions implied by an economic model. That actions reflect beliefs is, after all, a basic motivation for the literature on belief-driven fluctuations. Third, in a broad class of linear rational expectations models with unique equilibria, endogenous processes are purely a function of current and past fundamentals and beliefs about future fundamentals. So observational equivalence of fundamentals and beliefs implies observational equivalence of the entire economy.

We would also like to emphasize that the observability of beliefs distinguishes our concept of observational equivalence from that often encountered in time series analysis. To use a familiar example (cf. Hamilton, 1994, pp. 64-67), it is well-known that

$$y_t = \epsilon_t - \theta\epsilon_{t-1} \quad \epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2) \quad \text{and} \quad y_t = \tilde{\epsilon}_t - \tilde{\theta}\tilde{\epsilon}_{t-1} \quad \tilde{\epsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \tilde{\sigma}^2) \tag{3}$$

are two observationally equivalent representations of the stationary MA(1) process $\{y_t\}$ when $\tilde{\theta} = 1/\theta$ and $\sigma^2 = \theta^2\tilde{\sigma}^2$. However, this applies only when $\{y_t\}$ is the sole observable. If (rational) expectations of future values of $\{y_t\}$ are also observable, then the two representations in (3) are no longer the same. To see why, note that the variance of the one-step-ahead rational forecast $\hat{y}_t \equiv E_t[y_{t+1}]$ is equal to $\theta^2\sigma^2$ under the first representation, but σ^2 under the second. Therefore, an econometrician observing $\{\hat{y}_t\}$ *and* $\{y_t\}$ (or independent functions of these objects) could easily discriminate between these two representations.

The following proposition states the equivalence result for the simple example of this subsection, and provides the parametric mapping from one representation to the other. Its proof is in the Appendix.

Proposition 1. *The news representation* (1) *is observationally equivalent to the noise representation* (2) *if and only if:*

$$\sigma_x^2 = \sigma_{a,0}^2 + \sigma_{a,1}^2 \quad \textit{and} \quad \frac{\sigma_v^2}{\sigma_x^2} = \frac{\sigma_{a,0}^2}{\sigma_{a,1}^2}.$$

The intuition behind the result comes from the fact that the noise representation implies an observationally equivalent innovations representation (cf. Anderson and Moore, 1979, ch.

9) of the form:

$$x_t = \hat{x}_{t-1} + w_{0,t} \qquad \begin{bmatrix} w_{0,t} \\ w_{1,t} \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \kappa\sigma_v^2 & 0 \\ 0 & \sigma_x^2 + \sigma_v^2 \end{bmatrix}\right), \tag{4}$$
$$\hat{x}_t = \kappa w_{1,t}$$

where $\kappa = \sigma_x^2/(\sigma_x^2 + \sigma_v^2)$ is a Kalman gain parameter controlling how much people trust the noisy signal, and $w_t \equiv (w_{0,t}, w_{1,t})'$ is the vector of Wold innovations. But system (4) is the same as the news representation in system (1) when $a_{0,t} = w_{0,t}$ and $a_{1,t} = \kappa w_{1,t}$. The news shocks are linear combinations of the Wold innovations.

A direct implication of Proposition (1) is that the news representation is identified if and only if the noise representation is identified. By observational equivalence, both representations have the same likelihood function. Because the relations in Proposition (1) define a bijection, it is always possible to go from one set of parameters to the other and vice versa. This suggests that the distinction often made between news and noise representations in the structural VAR literature may be misleading. We take up that question in Section (4).

Proposition (1) also reveals that noise shocks are closely related to a popular thought experiment in the news-shock literature, which some researchers have used to isolate the effects of a change in beliefs that does not correspond to any change in fundamentals (e.g. Christiano et al. (2010) Section 4.2, Schmitt-Grohé and Uribe (2012) Section 4.2, Barsky, Basu, and Lee (2015) Section IV.A, or Sims (2016) Section 3.3). This experiment involves computing the impulse responses of endogenous variables to a current news shock followed by an offsetting future surprise shock.

In this simple example, it is easy to see that the noise shocks generate exactly the sort of offsetting news shocks envisioned by this thought experiment. Using the Kalman filter, the surprise and news shocks can be expressed as functions of fundamental and noise shocks:

$$a_{1,t} = \kappa x_{t+1} + \kappa v_t \quad \text{and} \quad a_{0,t} = (1-\kappa)x_t - \kappa v_{t-1}.$$

Therefore, a positive noise shock at date t generates a positive news shock at date t and an exactly offsetting surprise shock at date $t+1$.

This example shows that it may be possible to mimic noise shocks using particular linear combinations of news shocks. Nevertheless, there are a number of advantages to working directly with noise shocks. First, we can think about how often these situations arise, since we have an explicit probability distribution for the noise shocks: for example, how big is a "one standard deviation impulse" of a news reversal? Second, we can ask how important these types of news reversals are in the data overall; that is, we can do a proper variance decomposition. Third, in models with news shocks that are not i.i.d., it is not as straightforward to determine the configuration of news shocks that correspond to a noise shock. Therefore, it is desirable to have a more general characterization of the link between news and noise shocks. We turn to this more general characterization next.

2.2 Representation Theorem

This subsection generalizes the previous example to allow for news and noise at multiple future horizons, and potentially more complex time-series dynamics. To fix notation, we use $\mathcal{L}^2$ to denote the space of (equivalence classes of) random variables with finite second moments, which is a Hilbert space when equipped with the inner product $\langle a, b \rangle = E[ab]$ for any $a, b \in \mathcal{L}^2$. Completeness of this space is with respect to the norm $\|a\| \equiv \langle a, a \rangle^{1/2}$. For any collection of random variables in $\mathcal{L}^2$,

$$\{y_{i,t}\}, \quad \text{with } i \in \mathcal{I} \subseteq \mathbb{Z} \text{ and } t \in \mathbb{Z},$$

we let $\mathcal{H}_t(y)$ denote the closed subspace spanned by the variables $y_{i,\tau}$ for all $i \in \mathcal{I}$ and $\tau \in \mathbb{Z}$ such that $\tau \leq t$. To simplify notation, we write $\mathcal{H}(y) \equiv \mathcal{H}_\infty(y)$.

Fundamentals are summarized by a scalar discrete-time process $\{x_t\}$. As in the previous subsection, this process is taken to be mean-zero, stationary, and Gaussian.[8] The fact that $\{x_t\}$ is a scalar process is not restrictive; we can imagine a number of different scalar processes, each capturing changes in one particular fundamental. In that case it will be possible to apply the results from this section to each fundamental one at a time.

People's beliefs about fundamentals are summarized by a collection of random variables $\{\hat{x}_{i,t}\}$, with $i, t \in \mathbb{Z}$, where $\hat{x}_{i,t}$ represents the forecast of the fundamental realization x_{t+i} as of time t. Under the assumption of rational expectations, which is maintained throughout this paper, $\hat{x}_{i,t}$ is equal to the mathematical expectation of x_{t+i} with respect to a particular date-t information set. This, together with the restriction that all random variables generating the date-t information set are Gaussian, implies that the collection $\{\hat{x}_{i,t}\}$ fully characterizes people's entire subjective distribution over realizations of the sequence $\{x_t\}$.

A "representation of fundamentals and beliefs" means a specification of the fundamental process $\{x_t\}$ and the collection of people's conditional expectations about that process at each point in time $\{\hat{x}_{i,t}\}$. A typical assumption is that people's information set is equal to $\mathcal{H}_t(x)$, so $\hat{x}_{i,t} \in \mathcal{H}_t(x)$ for all $t \in \mathbb{Z}$. In this case, the process $\{x_t\}$ is itself sufficient to

[8]All our results can be extended to processes that are stationary only after suitable differencing. For example, if $\{x_t\}$ is not stationary, but is stationary after applying the filter $(1-L)^p$ for some $p > 0$, we can define $(1-\lambda L)^p \tilde{x}_t = (1-L)^p x_t$ for $0 < \lambda < 1$. Then $\{\tilde{x}_t\}$ is stationary, and we can apply our results to it and then let λ tend to 1 from below.

describe both the fundamental and people's beliefs about it. A key departure in models of belief-driven fluctuations is that people may have more information than what is reflected in $\mathcal{H}(x)$ alone; as a result, $\mathcal{H}(x) \subset \mathcal{H}(\hat{x})$. We therefore maintain this assumption throughout the paper. We also work exclusively with regular processes, in the sense of Rozanov (1967).

Definition 1. In a "news representation" of fundamentals and beliefs, the process $\{x_t\}$ is related to a collection of independent, stationary Gaussian processes $\{a_{i,t}\}$ with $i \in \mathcal{I} \subseteq \mathbb{Z}_+$ by the summation

$$x_t = \sum_{i \in \mathcal{I}} a_{i,t-i} \quad \text{for all } t \in \mathbb{Z},$$

where people's date-t information set is $\mathcal{H}_t(a) \supset \mathcal{H}_t(x)$.

The idea behind this representation is that people observe parts of the fundamental realization x_t prior to date t. The variable $\epsilon^a_{i,t} \equiv a_{i,t} - E[a_{i,t}|\mathcal{H}_{t-1}(a)]$ is called the "news shock" associated with horizon i whenever $i > 0$. By convention, $0 \in \mathcal{I}$, and in that case, the variable $\epsilon^a_{0,t}$ is referred to as the "surprise shock." An important aspect of this definition is that all of the news shocks are correlated both with fundamentals and people's beliefs. This is because any increase in fundamentals that people observe in advance must, other things equal, generate a one-for-one increase in fundamentals at some point in the future.

Definition 2. In a "noise representation" of fundamentals and beliefs, there is a collection of signal processes $\{s_{i,t}\}$ with $i \in \mathcal{I} \subseteq \mathbb{Z}_+$ of the form:[9]

$$s_{i,t} = m_{i,t} + v_{i,t}, \quad \text{for all } t \in \mathbb{Z},$$

where $m_{i,t} \in \mathcal{H}(x)$, $v_{i,t} \perp \mathcal{H}(x)$, and people's date-$t$ information set is $\mathcal{H}_t(s) \supset \mathcal{H}_t(x)$, which satisfies $\mathcal{H}_t(s) = \mathcal{H}_t(\hat{x})$.

The idea behind this representation is that people may receive signals about the fundamental realization x_t prior to date t, but those signals are contaminated with noise. The variable $\epsilon^v_{i,t} \equiv v_{i,t} - E[v_{i,t}|\mathcal{H}_{t-1}(v)]$ is called the "noise shock" associated with signal i. The variable $\epsilon^x_t \equiv x_t - E[x_t|\mathcal{H}_{t-1}(x)]$ is called the "fundamental shock." An important aspect of this definition is that all of the noise shocks are completely independent of fundamentals, but because people cannot separately observe $m_{i,t}$ and $v_{i,t}$ at date t, their beliefs are still affected by noise. The condition that $\mathcal{H}_t(s) = \mathcal{H}_t(\hat{x})$ simply rules out redundant or totally uninformative signals.

With these definitions in hand, we are ready to state the main result of the paper. Its proof is in the Appendix.

[9] In general, the index set $\mathcal{I}$ may be different from the one in the Definition (1).

Theorem 1. *Fundamentals and beliefs always have both a news representation and a noise representation. Moreover, the news representation is unique.*

This theorem clarifies the sense in which news and noise representations of fundamentals and beliefs are really just two sides of the same coin. It is possible to view the same set of data from either perspective. The proof is constructive, which means that it also provides an explicit computational method for passing from one representation to the other.

The only asymmetric aspect of the theorem involves the uniqueness of the two representations. Any particular news representation will be compatible with several different noise representations. This is the same sort of asymmetry present between signal models representations and innovations representations in the literature on state-space models. In general there exist infinitely many signal models with the same innovations representation. We explain in the subsequent sections, however, that this multiplicity of noise representations does not pose much of a problem.

An implication of Theorem (1) is that any model *economy* with a news representation of fundamentals and beliefs has an observationally equivalent version with a noise representation of fundamentals and beliefs, and vice versa. This is because the equivalence of fundamentals and beliefs implies the equivalence of any endogenous processes that are functions of them. To make this statement more precise, we first define here what we mean by an endogenous process, and then present this statement as a proposition. The proof of the proposition, together with all remaining proofs, are contained in the Online Appendix.

Definition 3. Given a fundamental process $\{x_t\}$ and a collection of forecasts $\{\hat{x}_{i,t}\}$ satisfying $\mathcal{H}_t(x) \subset \mathcal{H}_t(\hat{x})$, a process $\{c_t\}$ is "endogenous" with respect to $\{x_t\}$ if

$$c_t \in \mathcal{H}_t(\hat{x}) \quad \text{for all } t \in \mathbb{Z}.$$

Proposition 2. *If two different representations of fundamentals and beliefs are observationally equivalent, then they imply observationally equivalent dynamics for any endogenous process.*

The stipulation in Definition (3) that endogenous processes be linearly related to people's forecasts of fundamentals is not restrictive. Proposition (2) holds even if we generalize the definition of an endogenous process $\{c_t\}$ to require only that c_t be measurable with respect to people's date-t information set for all $t \in \mathbb{Z}$ (the proof provided in the Online Appendix establishes this stronger result). Together with Theorem (1), this means that as long as the random variables generating people's information set are jointly Gaussian, any *non-linear* economy that allows for belief-driven fluctuations can be equivalently written with either news or noise shocks.

Throughout the rest of the paper, however, we will retain the restriction of linearity in Definition (3). This is because the definitions of many objects of economic interest, such as variance decompositions, are typically defined only for linear models. Therefore, it is most natural to present the results in Sections (3) and (4) in terms of endogenous variables that can be expressed as linear functions of people's forecasts. Furthermore, all of the quantitative models we consider in Section (3.4) rely on linear-approximate equilibrium dynamics.

3 The Importance of Beliefs

A central question in the literature on belief-driven fluctuations is: how important are beliefs? Or, more precisely, how important are beliefs *relative* to fundamentals? Perhaps surprisingly, it turns out that no existing quantitative study in this literature has actually answered that question. Some studies report the importance of news shocks, which combine the contribution due to fundamentals with the contribution due independently to beliefs. Others include noise shocks and news shocks in the same model, and as a result, do not report the full importance of either one. In this section we argue that Theorem (1) provides a way to determine the importance of beliefs as an independent driver of fluctuations.

The first subsection explains the problem with using news shocks to determine the importance of beliefs, and the second subsection clarifies the problems that arise when attempting to include both news and noise shocks in the same model. To keep things clear, the discussion of both of these issues is framed in terms of the simple example from Section (2.1). The third subsection establishes a result regarding the uniqueness of variance decompositions.

3.1 The Problem with News Shocks

In the context of dynamic linear models, the importance of a set of exogenous shocks can be determined by performing a variance decomposition. This entails computing the model-implied variance of an endogenous process under the assumption that all shocks other than those in the set of interest are counterfactually equal to zero almost surely, and comparing that variance to the unconditional variance of the process. More nuanced versions include only considering variation over a certain range of spectral frequencies, or variation in forecast errors over a certain forecast horizon.

The problem with using news shocks to determine the importance of beliefs is that news shocks mix changes that are due to fundamentals and changes that are independently due to beliefs. This is because a news shock is an anticipated change in fundamentals. Expectations change at the time the news shock is realized, but then fundamentals change in the future when the anticipated change actually occurs. Of course, people's expectations may not always be fully borne out in future fundamental realizations, due to other unforeseen disturbances. Nevertheless, the anticipated shock is borne out on average, which is to say that news shocks are related to future fundamentals on average.

A stark way to see this point is to consider the importance of beliefs for driving fundamentals. Because fundamentals are purely exogenous, they are not driven by beliefs at all. However, in the simple news representation from Section (2.1), for example, news shocks can be an arbitrarily large part of fluctuations in the fundamental process $\{x_t\}$. Recall that in that example, $x_t = a_{0,t} + a_{1,t-1}$. Therefore, the fraction of the variation in $\{x_t\}$ due to news shocks, $\{a_{1,t}\}$ is given by:

$$\frac{\text{var}[x_t | a_{0,t} = 0]}{\text{var}[x_t]} = \frac{\text{var}[a_{1,t}]}{\text{var}[x_t]} = \frac{\sigma^2_{a,1}}{\sigma^2_{a,0} + \sigma^2_{a,1}}.$$

As $\sigma^2_{a,1}$ increases relative to $\sigma^2_{a,0}$, this fraction approaches one, in which case news shocks would explain all the variation in $\{x_t\}$.

To disentangle the importance of beliefs from fundamentals in models with news shocks, we can use Theorem (1). Specifically, we can write down an observationally equivalent noise representation of the news model, and then use a variance decomposition to compute the share of variation attributable to noise shocks. Because these shocks are unrelated to fundamentals at all horizons, they capture precisely the independent contribution of beliefs.

Returning to the example from Section (2.1), we have already shown that an observationally equivalent noise representation involves $x_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_x)$ with $\sigma^2_x \equiv \sigma^2_{a,0} + \sigma^2_{a,1}$. Therefore, the fraction of variation in $\{x_t\}$ due to noise shocks is:

$$\frac{\text{var}[x_t | x_t = 0]}{\text{var}[x_t]} = 0,$$

which is the correct answer to the question of how much beliefs contribute to the fluctuations of fundamentals. This example illustrates the more general point that in order to determine the importance of beliefs, one should perform variance decompositions in terms of noise shocks rather than news shocks.

The fact that variance decompositions in terms of news shocks are not appropriate for determining the importance of beliefs has lead some researchers to conclude that there is a fundamental problem with using variance decompositions for that purpose.[10] We would

[10] For example, Sims (2016) p.42 describes the problem of identifying the importance of beliefs (which both he and Barsky et al. (2015) call "pure news") as a fundamental limitation of the traditional variance decomposition.

like to suggest that the problem is not with variance decompositions as such; rather, the problem is with the type of shock one considers. It is noise shocks, not news shocks, that are the appropriate shocks for isolating the independent contribution of beliefs. Once that distinction has been made, traditional variance decompositions can be performed as usual.

3.2 Mixing News and Noise Shocks

In some cases, researchers have constructed representations of fundamentals and beliefs that seem to include both news and noise shocks at the same time (e.g. Blanchard et al., 2013; Barsky and Sims, 2012). A simple example is:

$$\begin{aligned} x_t &= \lambda_{t-1} + \eta_t \\ s_t &= \lambda_t + \xi_t \end{aligned} \qquad \begin{bmatrix} \eta_t \\ \lambda_t \\ \xi_t \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & \sigma_\lambda^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix}\right). \tag{5}$$

At each date t, people observe $\{x_\tau, s_\tau\}$ for all $\tau \leq t$. The shock λ_t looks like a news shock because it affects people's beliefs at date t (through the signal s_t), but does not affect fundamentals until the following period. Similarly, the shock η_t looks like a surprise shock because it affects people's beliefs and the fundamental at the same time. Finally, the shock ξ_t looks like a noise shock because it affects people's beliefs but is independent of fundamentals.

The problem with this type of representation, at least from the perspective of isolating the importance of beliefs, is that while ξ_t is unrelated to fundamentals, it does not fully capture the independent contribution of beliefs. This is because λ_t and η_t are not purely functions of fundamental shocks; they also capture changes that are independently due to beliefs. To see this, notice that in the limit case $\xi_t = 0$, we have that $s_t = \lambda_t$ and this representation collapses to a news representation with $a_{0,t} \equiv \eta_t$ and $a_{1,t} \equiv \lambda_t$. We have already seen in Proposition (1) that such a news representation has an observationally equivalent noise representation with (non-zero) noise shocks. Therefore $\xi_t = 0$ does not mean that beliefs do not have an independent role to play as a driver of fluctuations.

Of course, Theorem (1) implies that the representation in (5), which is neither news or noise representation, still has an observationally equivalent noise representation. The following proposition presents the mapping from one representation to the other.

Proposition 3. *The representation of fundamentals and beliefs in* (5) *is observationally equivalent to the noise representation in* (2) *if and only if:*

$$\sigma_x^2 = \sigma_\lambda^2 + \sigma_\eta^2 \quad \textit{and} \quad \frac{\sigma_v^2}{\sigma_x^2} = \frac{\sigma_\lambda^2(\sigma_\eta^2 + \sigma_\xi^2) + \sigma_\eta^2 \sigma_\xi^2}{\sigma_\lambda^4}.$$

To see how the process $\{\xi_t\}$ understates the importance of beliefs, consider the endogenous variable $\hat{x}_t = E_t[x_{t+1}]$. Under representation (5), $\hat{x}_t = \frac{\sigma_\lambda^2}{\sigma_\lambda^2+\sigma_\xi^2}(\lambda_t + \xi_t)$, so the contribution of the process $\{\xi_t\}$ is

$$\frac{\text{var}[\hat{x}_t|\lambda_t = \eta_t = 0]}{\text{var}[\hat{x}_t]} = \frac{\sigma_\xi^2}{\sigma_\lambda^2 + \sigma_\xi^2}.$$

On the other hand, in the observationally equivalent noise representation implied by Proposition (3), $\hat{x}_t = \frac{\sigma_x^2}{\sigma_x^2+\sigma_v^2}(x_{t+1} + v_t)$. Therefore, the contribution of $\{v_t\}$ is

$$\frac{\text{var}[\hat{x}_t|x_t = 0]}{\text{var}[\hat{x}_t]} = \frac{\sigma_v^2}{\sigma_x^2 + \sigma_v^2} = \frac{\sigma_\lambda^2\sigma_\eta^2}{\sigma_\lambda^2 + \sigma_\eta^2} + \frac{\sigma_\xi^2}{\sigma_\lambda^2 + \sigma_\xi^2},$$

where the second equality uses the parametric restrictions from Proposition (3). Because the first term in this expression is positive, it follows that $\{\xi_t\}$ understates the importance of beliefs for explaining variations in $\{\hat{x}_t\}$. It is also easy to see how the importance of beliefs can be strictly positive even as $\sigma_\xi^2 \to 0$.

3.3 Different Noise Representations

So far we have argued that it is possible to use a noise representation to determine the importance of beliefs as an independent driver of fluctuations. First, one can rewrite any representation of fundamentals and beliefs as a noise representation using the constructive procedure from Theorem (1). Then, one can use a variance decomposition to determine the share of variation in any endogenous variable that is attributable to noise shocks. And this share represents the independent contribution of beliefs.

But is the variance decomposition in terms of noise shocks unique? As we pointed out in the discussion of Theorem (1), any representation of fundamentals and beliefs is compatible with infinitely many different noise representations. Fortunately, it turns out that all observationally equivalent noise representations deliver the same answer regarding the importance of beliefs for any endogenous process. For variance decompositions, the fact that noise representations are not unique is not a problem.

Proposition 4. *In any noise representation of fundamentals and beliefs, the variance decomposition of any endogenous process in terms of noise and fundamentals is uniquely determined over any frequency range.*

An immediate corollary of this proposition is that the variance decomposition of people's errors in forecasting an endogenous process is also uniquely determined for any forecast horizon. This is because the forecast errors are themselves endogenous processes to which Proposition (4) applies.

Corollary 1. *In any noise representation of fundamentals and beliefs, the forecast error variance decomposition of any endogenous process in terms of noise and fundamentals is uniquely determined for any horizon, and over any frequency range.*

3.4 Quantifying the Importance of Beliefs

In this subsection, we use Theorem (1) and Proposition (4) to empirically quantify the independent contribution of beliefs in driving business-cycle fluctuations. Because several models of belief-driven fluctuations have already been constructed and estimated in the literature, we take something of a meta-analytic perspective. Specifically, we select three prominent theoretical models that have been estimated in the literature and compute the importance of beliefs implied by each of those models for different macroeconomic variables (e.g. output, investment, etc.). The three models are: the model of news shocks from Schmitt-Grohé and Uribe (2012), the model of news and animal spirits from Barsky and Sims (2012), and the model of noise shocks from Blanchard et al. (2013).

These three models are different in several respects. First, they incorporate different physical environments, including differences in preferences, frictions and market structure. Second, the three models are estimated on different data and with different sample periods. Third, the authors make different assumptions about the information structure faced by agents. While agents in all three models observe current fundamentals and receive advance information about future fundamentals, Schmitt-Grohé and Uribe (2012) take a pure news perspective while the Barsky and Sims (2012) and Blanchard et al. (2013) offer somewhat different perspectives on combining news and noise within a single model.

Perhaps not surprisingly given the scope of these differences, the authors above come to very different conclusions. Schmitt-Grohé and Uribe (2012) conclude that news shocks explain about one half of aggregate fluctuations, but do not take an explicit stance on the importance of independent fluctuations in beliefs. Barsky and Sims (2012) also conclude that news shocks are important, and that noise shocks explain essentially none of the variation in any variable. However, Blanchard et al. (2013) conclude that noise shocks play a crucial role in business cycle dynamics, especially for consumption.

In principle, it is possible that these different conclusions are largely a result of the different "normalizations" the authors take with respect to noise shocks. Indeed, our analysis indicates that all authors have (implicitly or explicitly) underestimated the actual of role of independent shocks to beliefs in their estimated economies. For Schmitt-Grohé and Uribe (2012) and Barsky and Sims (2012), we find that the role of noise rises from being essentially zero to being small but non-trivial, generally between 3% and 11% at the business cycle frequency. Surprisingly, even Blanchard et al. (2013) significantly underestimate the role of pure noise in driving their economy, with beliefs about productivity driving endogenous variables more than productivity itself. While our results indicate that noise shocks are more important than previously reported, they do not fully explain the degree of disagreement regarding the independent contribution of beliefs.

3.4.1 Schmitt-Grohé and Uribe (2012)

The first model comes from Schmitt-Grohé and Uribe (2012), and was constructed to determine the importance of news shocks for explaining aggregate fluctuations in output, consumption, investment, and employment. The main result of their paper is that news shocks account for about half of the predicted aggregate fluctuations in those four variables. As we have seen in the previous section, however, news shocks mix fluctuations due to beliefs and fundamentals. As a result, exactly what this model implies about the importance of beliefs is still an unanswered question.

The model is a standard real business cycle model with six modifications: investment adjustment costs, variable capacity utilization with respect to the capital stock, decreasing returns to scale in production, one period internal habit formation in consumption, imperfect competition in labor markets, and period utility allowing for a low wealth effect on labor supply. Fundamentals comprise seven different independent processes, which capture exogenous variation in: stationary and non-stationary neutral productivity, stationary and non-stationary investment-specific productivity, government spending, wage markups, and preferences. The model is presented in more detail in the Online Appendix (B.1).

Each of the seven exogenous fundamentals follows a law of motion:

$$x_t = \rho_x x_{t-1} + \epsilon^a_{0,t} + \epsilon^a_{4,t-4} + \epsilon^a_{8,t-8}, \qquad \begin{bmatrix} \epsilon^a_{0,t} \\ \epsilon^a_{4,t} \\ \epsilon^a_{8,t} \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma^2_{a,0} & 0 & 0 \\ 0 & \sigma^2_{a,4} & 0 \\ 0 & 0 & \sigma^2_{a,8} \end{bmatrix}\right). \tag{6}$$

where $0 < \rho_x < 1$. The model is estimated by likelihood-based methods on a sample of quarterly U.S. data from 1955:Q2-2006:Q4. The time series used for estimation are: real GDP, real consumption, real investment, real government expenditure, hours, utilization-adjusted total factor productivity, and the relative price of investment.

A variance decomposition shows that news shocks turn out to be very important. The first column of Table (1) shows the share of business-cycle variation in the level of four endogenous variables that is attributable to surprise shocks $\{\epsilon^a_{0,t}\}$, and the second column shows the share attributable to the news shocks $\{\epsilon^a_{4,t}\}$ and $\{\epsilon^a_{8,t}\}$ combined. We define business cycle frequencies as the components of the endogenous process with periods of 6 to 32 quarters, and we focus on variance decompositions over these frequencies to facilitate comparison across the different models in this section. Our results are consistent with the authors' original findings (see their Table V).

However, to determine the contribution of beliefs relative to fundamentals, we would like to construct a noise representation that is observationally equivalent to representation (6). One such noise representation is in the following proposition.

Proposition 5. *The representation of fundamentals and beliefs in system* (6) *is observationally equivalent to the noise representation*

$$\begin{aligned} x_t &= \rho_x x_{t-1} + \epsilon_t^x \\ s_{4,t} &= \epsilon_{t+4}^x + v_{4,t} \\ s_{8,t} &= \epsilon_{t+8}^x + v_{8,t}, \end{aligned} \qquad \begin{bmatrix} \epsilon_t^x \\ v_{4,t} \\ v_{8,t} \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_{v,4}^2 & 0 \\ 0 & 0 & \sigma_{v,8}^2 \end{bmatrix}\right)$$

with the convention that $s_{0,t} \equiv x_t$*, and where*

$$\begin{aligned} \sigma_x^2 &= \sigma_{a,0}^2 + \sigma_{a,4}^2 + \sigma_{a,8}^2 \\ \sigma_{v,4}^2 &= \frac{1}{\sigma_{a,4}^2}\sigma_{a,0}^2(\sigma_{a,0}^2 + \sigma_{a,4}^2) \\ \sigma_{v,8}^2 &= \frac{1}{\sigma_{a,8}^2}(\sigma_{a,0}^2 + \sigma_{a,4}^2)(\sigma_{a,0}^2 + \sigma_{a,4}^2 + \sigma_{a,8}^2). \end{aligned}$$

We can use the noise representation in Corollary (5) with the same parameter estimates as before, and re-compute the variance decomposition of the seven observable variables in terms of fundamental shocks and noise shocks. This decomposition is unique by Proposition (4). There is no need to re-estimate the model because observational equivalence implies that the likelihood function is the same under both representations. The third column of Table (1) shows the share of variation attributable to fundamental shocks $\{\epsilon_t^x\}$, and the fourth column shows the share attributable to the noise shocks $\{v_{4,t}\}$ and $\{v_{8,t}\}$ combined.

The main result is that nearly all of the variation in output, consumption, investment, and hours is due to fundamentals. In terms of differences across the endogenous variables, it is interesting that real investment growth is affected the least by news shocks, but it is affected the most by noise shocks. At the same time, hours worked is affected the most by news shocks and the least by noise shocks. But based on the fact that 89% or more of the variation in every series is attributable to fundamental changes, we conclude that beliefs are not an important independent source of fluctuations through the lens of this model.

Variable	Surprise	News	Fundamental	Noise
Output	57	43	94	6
Consumption	50	50	95	5
Investment	55	45	89	11
Hours	16	84	97	3

Table 1: Variance decomposition () in the model of Schmitt-Grohé and Uribe (2012) over business cycle frequencies of 6 to 32 quarters. All variables are in levels. Estimated model parameters are set to their posterior median values.

3.4.2 Barsky and Sims (2012)

The second model comes from Barsky and Sims (2012). It was constructed to determine whether measures of consumer confidence change in ways that are related to macroeconomic aggregates because of noise (i.e. "animal spirits") or news. The main result of the paper is that changes in consumer confidence are mostly driven by news and not noise. Noise shocks account for negligible shares of the variation in forecast errors of consumption and output, while news shocks account for over half of the variation in long-horizon forecast errors. However, as we saw in Section (3.2), including both news and noise shocks in the same model can be problematic when it comes to isolating the importance of beliefs.

The model is a standard New-Keynesian DSGE model with real and nominal frictions: one period internal habit formation in consumption, capital adjustment costs (as opposed to investment adjustment costs, according to which costs are expressed as a function of the growth rate of investment rather than the level of investment relative to the existing capital stock), and monopolistic price setting with time-dependent price rigidity. Fundamentals comprise three different independent processes, which capture exogenous variation in: non-stationary neutral productivity, government spending, and monetary policy. The model is presented in more detail in Online Appendix (B.2).

People only receive advance information about productivity, and not about the other two fundamentals. So it is only beliefs about productivity that can play an independent role in driving fluctuations. Letting x_t denote the growth rate of productivity (in deviations from its mean), and using our notation from Section (3.2), the process $\{x_t\}$ is assumed to follow a law of motion of the form:

$$\begin{aligned} x_t &= \lambda_{t-1} + \eta_t \\ \lambda_t &= \rho\lambda_{t-1} + \epsilon_t^\lambda \\ s_t &= \lambda_t + \xi_t \end{aligned} \qquad \begin{bmatrix} \epsilon_t^\lambda \\ \eta_t \\ \xi_t \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_\lambda^2 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix}\right), \tag{7}$$

where $0 < \rho < 1$. Barsky and Sims (2012) refer to ϵ_t^λ as a news shock, η_t as a surprise shock, and ξ_t as a noise (animal spirits) shock.[11] However, these definitions are not consistent with the definitions in our paper. To avoid any confusion we will use asterisks to indicate the terminology of Barsky and Sims (2012). So we refer to ϵ_t^λ is a news* shock, η_t is a surprise* shock, and ξ_t is a noise* shock.

The model is estimated by minimizing the distance between impulse responses generated from simulations of the model and those from estimated structural vector autoregressions. The vector autoregressions are estimated on quarterly U.S. data from 1960:Q1-2008:Q4. The time series used to estimate the vector autoregression are: real GDP, real consumption, CPI inflation, a measure of the real interest rate, and a measure of consumer confidence from the Michigan Survey of Consumers (E5Y).

A variance decomposition shows that news* shocks are much more important than noise* shocks. The first column of Table (2) shows the share of business-cycle variation in the level of four endogenous variables that is attributable to surprise* shocks $\{\eta_t\}$, the second shows the share attributable to news* shocks $\{\epsilon_t^\lambda\}$, and the third shows the share attributable to noise* shocks $\{\xi_t\}$. Due to the presence of exogenous government spending and monetary policy shocks, the rows do not sum to 100%; the residual represents the combined contribution of these two additional fundamental shocks. These results are consistent with the authors' original findings, which are stated in terms of the variance decompositions of forecast errors over different horizons, but across all frequency ranges (see their Table 3).

To properly isolate the independent contributions of beliefs, we would again like to construct a noise representation that is observationally equivalent to representation (7). The following proposition presents one such noise representation.

Proposition 6. *The representation of fundamentals and beliefs in system* (7) *is observationally equivalent to the noise representation*

$$\begin{aligned} x_t &= -\rho \frac{\sigma_\eta^2}{\sigma_\lambda^2}\left[m_t - \left(\frac{1+\delta^2}{\delta}\right) m_{t-1} + m_{t-2}\right] \\ m_t &= (\rho+\delta) m_{t-1} - \rho\delta m_{t-2} + \epsilon_t^m \\ s_t &= m_t + v_t \\ v_t &= \delta v_{t-1} + \epsilon_t^v - \beta \epsilon_{t-1}^v \end{aligned}$$

$$\begin{bmatrix} \epsilon_t^m \\ \epsilon_t^v \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \delta\sigma_\lambda^4/(\rho\sigma_\eta^2) & 0 \\ 0 & \delta\sigma_\xi^2/\beta \end{bmatrix}\right),$$

[11] While these authors refer to signal noise as "animal spirits," they also use the term "pure noise" to refer to statistical measurement error. We are only concerned with noise in the first sense.

with the convention that $s_{0,t} \equiv x_t$*, and where*

$$\delta = \frac{1}{2\rho}\left(1+\rho^2+\frac{\sigma_\lambda^2}{\sigma_\eta^2}-\left[\left(1+\rho^2+\frac{\sigma_\lambda^2}{\sigma_\eta^2}\right)^2-4\rho^2\right]^{1/2}\right)$$

$$\beta = \frac{1}{2\rho}\left(1+\rho^2+\frac{\sigma_\lambda^2(\sigma_\lambda^2+\sigma_\xi^2)}{\sigma_\eta^2\sigma_\xi^2}-\left[\left(1+\rho^2+\frac{\sigma_\lambda^2(\sigma_\lambda^2+\sigma_\xi^2)}{\sigma_\eta^2\sigma_\xi^2}\right)^2-4\rho^2\right]^{1/2}\right).$$

Using the noise representation in this proposition, we can re-compute the variance decomposition of the endogenous processes in terms of fundamental shocks and noise shocks. The fourth column of Table (2) shows the share of variation attributable to fundamental productivity shocks, and the fifth column shows the share attributable to productivity noise shocks. Again, the rows do not sum to 100% due to the presence of government spending and monetary policy shocks. Conceptually, the contribution of these shocks should also be included under the heading of fundamental shocks, but for comparison with the first three columns, we only include fundamental productivity shocks in the fourth column.

Variable	Surprise*	News*	Noise*	Fundamental	Noise
Output	53	37	0	89	1
Consumption	61	34	1	89	9
Investment	40	43	1	80	4
Hours	62	14	0	75	3

Table 2: Variance decomposition (%) in the model of Barsky and Sims (2012) over business cycle frequencies of 6 to 32 quarters. All variables are in levels, and estimated parameters are set to their point-estimated values. The rows do not sum to 100% because of other non-technology fundamental processes. Asterisks refer to the authors' terminology.

As in the model of Schmitt-Grohé and Uribe (2012), nearly all of the variation in output, consumption, investment, and hours is due to fundamentals. The contribution of noise shocks is larger than the contribution of noise* shocks, for all variables. However, the bulk of the contribution of news* shocks turns out to be due to fundamentals rather than noise.

3.4.3 Blanchard, L'Huillier, and Lorenzoni (2013)

The third model we consider comes from Blanchard et al. (2013), and was constructed "to separate fluctuations due to changes in fundamentals (news) from those due to temporary errors in agents' estimates (noise)."[12] The main quantitative result of their paper is that

[12] This quotation is taken from the article's abstract, which can be found on the AEA's website: `https://www.aeaweb.org/articles?id=10.1257/aer.103.7.3045`.

noise shocks explain a sizable fraction of short-run consumption fluctuations. However, it turns out that what the authors call "noise" shocks do not fully isolate fluctuations due to temporary errors in agents' estimates. So it remains an unanswered question what this model implies about the importance of beliefs.

The model is a standard New Keynesian DSGE model with real and nominal frictions: one-period internal habit formation in consumption, investment adjustment costs, variable capital capacity utilization, and monopolistic price and wage setting with time-dependent price rigidities. Fundamentals comprise six different independent processes, which capture exogenous variation in: non-stationary neutral productivity, stationary investment-specific productivity, government spending, wage markups, final good price markups, and monetary policy. For more details, see Online Appendix (B.3).

People only receive advance information about productivity, and not about the other five fundamentals. So it is only beliefs about productivity that can play an independent role in driving fluctuations. Let x_t denote the level of productivity, which is observed by people in the economy, and let s_t denote the additional informative signal that people receive. Then the processes $\{s_t\}$ and $\{x_t\}$ are assumed to evolve according to a system of the form

$$\begin{aligned} x_t &= \lambda_t + \eta_t \\ s_t &= \lambda_t + \xi_t \\ \Delta\lambda_t &= \rho\Delta\lambda_{t-1} + \epsilon_t^\lambda \\ \eta_t &= \rho\eta_{t-1} + \epsilon_t^\eta \end{aligned} \qquad \begin{bmatrix} \epsilon_t^\lambda \\ \epsilon_t^\eta \\ \xi_t \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_\lambda^2 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix}\right), \tag{8}$$

with the parameter restriction that $\rho\sigma_\lambda^2 = (1-\rho)^2\sigma_\eta^2$.

The authors refer to ϵ_t^λ as a permanent productivity shock, ϵ_t^η as a transitory productivity shock, and ξ_t as a noise shock. Taken together, they refer to ϵ_t^λ and ϵ_t^η as news shocks, because they are both correlated with future productivity. Again, because these definitions are not consistent with the ones in our paper, we will use asterisks to indicate the authors' terminology in contrast to ours.

The model is estimated using likelihood-based methods on a sample of quarterly U.S. data from 1954:Q3-2011:Q1. The time series used for estimation are: real GDP, real consumption, real investment, employment, the federal funds rate, inflation as measured by the implicit GDP deflator, and wages.

A variance decomposition reveals that noise* shocks are important, especially for consumption. The first column of Table (3) shows the share of business-cycle variation in the level of output, consumption, investment, and hours that is attributable to news* shocks, $\{\epsilon_t^\lambda\}$ and $\{\epsilon_t^\eta\}$, and the second column shows the share attributable to noise* shocks $\{\xi_t\}$.

Due to the presence of the other five fundamental shocks, the rows do not sum to 100%; the residual represents the combined contribution of these additional fundamental shocks. These results are consistent with the authors' original findings, which are stated in terms of the variance decompositions of forecast errors over different horizons (see their Table 6).

However, to properly isolate the independent contribution of beliefs, we can derive a noise representation that is observationally equivalent to representation (8). The following proposition presents one such noise representation.

Proposition 7. *The representation of fundamentals and beliefs in system* (8) *is observationally equivalent to the noise representation*

$$\begin{aligned}
x_t &= -\frac{\rho}{(1-\rho)^2}m_{t+1} + \frac{(1+\rho^2)}{(1-\rho)^2}m_t - \frac{\rho}{(1-\rho)^2}m_{t-1} \\
s_t &= m_t + v_t \\
m_t &= (1+2\rho)m_{t-1} - \rho(2+\rho)m_{t-2} + \rho^2 m_{t-3} + \epsilon_t^m \\
v_t &= 2\rho v_{t-1} - \rho^2 v_{t-2} + \epsilon_t^v - (\delta+\bar{\delta})\epsilon_{t-1}^v + \delta\bar{\delta}\epsilon_{t-2}^v
\end{aligned}$$

$$\begin{bmatrix} \epsilon_t^m \\ \epsilon_t^v \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(0, \begin{bmatrix} (1-\rho)^2\sigma_\lambda^2 & 0 \\ 0 & \rho^2\sigma_\xi^2/(\delta\bar{\delta}) \end{bmatrix}\right),$$

with the convention that $s_{0,t} = x_t$, *and where*[13]

$$\delta = \frac{1}{2\rho}\left(1+\rho^2+\rho^{1/2}\frac{\sigma_\lambda}{\sigma_\xi}i - \left[\left(1+\rho^2+\rho^{1/2}\frac{\sigma_\lambda}{\sigma_\xi}i\right)^2 - 4\rho^2\right]^{1/2}\right).$$

Using the noise representation in this proposition, we can re-compute the variance decomposition of the endogenous processes in terms of fundamental shocks and noise shocks. The fourth column of Table (3) shows the share of variation attributable to fundamental productivity shocks and the fifth column shows the share attributable to productivity noise shocks. Again, the rows do not sum to 100% due to the presence of fundamental processes other than productivity.

In contrast to both the Schmitt-Grohé and Uribe (2012) and Barsky and Sims (2012) models, we find that a sizable fraction of the variation in output, consumption, and hours worked can be attributed to noise shocks. For example, nearly 60% of the variation in consumption is due to noise shocks. This is more than 10% larger than the share Blanchard et al. (2013) originally attributed to independent fluctuations in beliefs. A result of similar

[13]In the definition of δ, $i \equiv \sqrt{-1}$ is the imaginary unit, and $\bar{\delta}$ denotes the complex conjugate of δ. Both $\delta + \bar{\delta}$ and $\delta\bar{\delta}$ are real numbers.

Variable	News*	Noise*	Fundamental	Noise
Output	34	22	26	29
Consumption	40	44	27	57
Investment	6	3	4	5
Hours	17	29	7	39

Table 3: Variance decomposition (%) in the model of Blanchard et al. (2013) over business cycle frequencies of 6 to 32 quarters. All variables are in levels, and estimated parameters are set to their posterior median values. The rows do not sum to 100% because of other non-technology fundamental processes.

magnitude is true for output and hours worked. Moreover, it is interesting that for all variables in the table, noise about productivity is in fact more important than productivity itself. This cannot be seen from the original decomposition.

3.5 Past Versus Future Fundamentals

So far, our discussion has focused on the distinction between the relative contribution of fundamental and non-fundamental (noise) shocks. Another interesting distinction involves the relative importance of future fundamental shocks compared to current and past fundamental shocks. Even if noise shocks are irrelevant, future fundamental shocks could still be important factors determining current actions.

For example, in models in which people have perfect foresight, the contribution of noise shocks is exactly zero; but if the model is purely forward-looking, future fundamental shocks could explain all of the variation in current actions. Similarly, in models in which people receive very accurate (but not perfect) information about future fundamentals, noise shocks can matter very little while future fundamentals still matter very much.

In both the models of Schmitt-Grohé and Uribe (2012) and Barsky and Sims (2012), fundamental shocks explain the bulk of fluctuations. Is that because agents are correctly anticipating future fundamental changes before they occur, or because they are merely reacting to past fundamental changes? To answer this question, we need to perform a different sort of decomposition.

Letting $\mathcal{H}(s)$ denote the space spanned by signals in a noise representation, we can perform the unique decomposition

$$\mathcal{H}(s) = \Big[\ \underbrace{\mathcal{H}_t(\epsilon^x)}_{\text{past/present}} \ \oplus\ \underbrace{\mathcal{H}^{t+1}(\epsilon^x)}_{\text{future}}\ \Big] \ \oplus\ \underbrace{\mathcal{H}(\epsilon^v)}_{\text{noise}},$$

where $\mathcal{H}^{t+1}(\epsilon^x)$ denotes the closed subspace spanned by the fundamental shocks ϵ^x_τ for all

$\tau \geq t+1$. Because any endogenous processes $\{c_t\}$ satisfies $c_t \in \mathcal{H}(s)$ for all t, it follows that c_t can be uniquely decomposed into the sum of three independent random variables: a_t^p, which is an element of $\mathcal{H}_t(\epsilon^x)$ and captures the contribution of past and present fundamental shocks; a_t^f, which is an element of $\mathcal{H}^{t+1}(\epsilon^x)$ and captures the contribution of future fundamental shocks; and b_t, which is an element of $\mathcal{H}(\epsilon^v)$ and captures the contribution of noise shocks:

$$c_t = a_t^p + a_t^f + b_t. \tag{9}$$

Because $a_t^p \perp a_t^f \perp b_t$ for each t, $\text{var}[c_t] = \text{var}[a_t^p] + \text{var}[a_t^f] + \text{var}[b_t]$. Therefore, the fraction of variation in $\{c_t\}$ due to future fundamentals is equal to $\text{var}[a_t^f]/\text{var}[c_t]$.

One complication relative to the result in Proposition (4) is that this sort of past versus future decomposition does not make sense in the frequency domain. While $a_t^p \perp a_t^f$ for each t, it is not true that $a_t^p \perp a_s^f$ for all $t \neq s$. The distinction between past and future is different relative to different dates. One date's past shock is another date's future shock. Since every frequency ω contains contributions from all time periods, there is no sensible way to measure the contribution of future fundamental shocks "over business cycle frequencies." For the same reason, it would also be inappropriate to render difference-stationary processes stationary by applying a band-pass filter which isolates certain frequency ranges.

To separately compute the contributions of past and future fundamentals, we therefore use a flexible exponential de-trending procedure that preserves the distinction between past and future shocks. For a difference-stationary process $\{y_t\}$, we define the stochastic trend $\bar{y}_t$ to be an exponential moving average of past values:

$$\bar{y}_t = (1-\lambda)y_{t-1} + \lambda \bar{y}_{t-1},$$

where $0 \leq \lambda < 1$. We then define the de-trended process to be $\tilde{y}_t \equiv y_t - \bar{y}_t$.

The parameter λ controls the extent to which the trend depends on past values. When $\lambda = 0$, $\tilde{y}_t = \Delta y_t$, so the de-trended process is the first-differenced version of the original process. As $\lambda \to 1$, $\tilde{y}_t \to y_t$. By varying λ, we can therefore consider a range of different hypotheses regarding the stochastic trend. Because the filter is one-sided for any λ (unlike frequency-domain filters), it is always the case that $\mathcal{H}_t(y) = \mathcal{H}_t(\tilde{y})$ for each t, which means that the past and future in terms of $\tilde{y}_t$ is the same as the past and future in terms of y_t.[14]

Figure (1) plots the fraction of the fundamental share due to future fundamental shocks, for each of the three models considered in this section. In the notation of Equation (9), that means we plot the fraction $\text{var}[a_t^f]/(\text{var}[a_t^p] + \text{var}[a_t^f])$. We plot this fraction for a range of different de-trended versions of the endogenous variables, corresponding to a different values

[14] In this respect, our proposal is similar in spirit to the procedure recently suggested by Hamilton (2017).

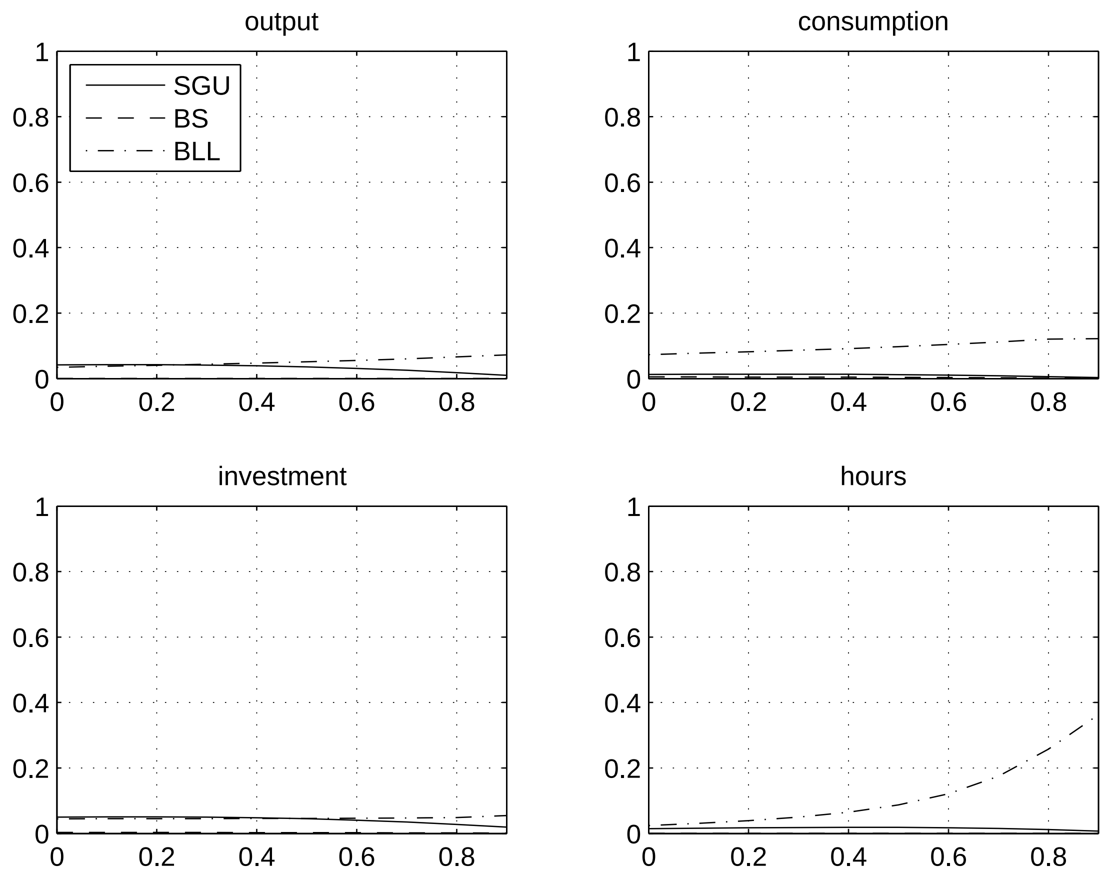

Figure 1: Fraction of the fundamental share due to future fundamental shocks, as a function of the de-trending parameter $0 \leq \lambda < 1$. $\lambda = 0$ corresponds to a decomposition in (log) first differences, and $\lambda \to 1$ corresponds to a decomposition in (log) levels.

of λ. As in the previous decompositions in this section, we focus only on fundamentals about which agents receive some advance information. That means that for the models of Barsky and Sims (2012) and Blanchard et al. (2013), we focus only on productivity, while in the model of Schmitt-Grohé and Uribe (2012) we include all seven fundamentals.[15]

The consistent result across all three models is that the bulk of the contribution of fundamentals comes from current and past — not future — fundamental shocks. In some cases, it is difficult to see that there are actually three lines in each subplot. This is because one of the lines is visually indistinguishable from zero. In the model of Barsky and Sims (2012), endogenous variables are the least sensitive to future shocks (on average across λ), followed by the model of Schmitt-Grohé and Uribe (2012) and then Blanchard et al. (2013).

This result may seem surprising considering that news shocks are fairly important in all three models. How can it be that news shocks are so important, but future fundamental

[15] If we include other non-productivity fundamentals in the first two models as well, the denominator of $\text{var}[a_t^f]/(\text{var}[a_t^p] + \text{var}[a_t^f])$ increases, implying an even smaller relative contribution of future fundamentals.

shocks are not? The answer is that news shocks not only mix fluctuations due to fundamentals and noise (as we discuss in Section (3.1)), they *also* mix fluctuations due to past and future fundamentals.

More precisely, for future fundamental shocks to matter, it is necessary but not sufficient that news shocks matter. In models with no news shocks, current outcomes must always be a function of current and past fundamental shocks, so future fundamental shocks are irrelevant. In models with news shocks, current outcomes can depend on future fundamentals. However, if the model is not sufficiently "forward-looking" (at least with respect to certain endogenous variables), it may well be the case that future fundamentals are still unimportant. However, because news shocks eventually show up as changes in fundamentals, current actions may nevertheless depend heavily on "past" news shocks.

To see how future fundamentals can matter very little, while news shocks can matter very much, recall the i.i.d. news model from Section (2.1):

$$x_t = a_{0,t} + a_{1,t-1}, \quad \begin{bmatrix} a_{0,t} \\ a_{1,t} \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_{a,0}^2 & 0 \\ 0 & \sigma_{a,1}^2 \end{bmatrix}\right).$$

Consider an endogenous variable that depends on current and expected future fundamentals with weights ϕ_0 and ϕ_1:

$$c_t = \phi_0 x_t + \phi_1 E_t[x_{t+1}]. \tag{10}$$

Writing the expectation as a function of future fundamental shocks, we have $c_t = \phi_0 x_t + \phi_1 \kappa (x_{t+1} + v_t)$, where $\kappa \equiv \sigma_{a,1}^2 / (\sigma_{a,0}^2 + \sigma_{a,1}^2)$ and

$$\begin{bmatrix} x_t \\ v_t \end{bmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \begin{bmatrix} \sigma_{a,0}^2 + \sigma_{a,1}^2 & 0 \\ 0 & (\sigma_{a,0}^2 + \sigma_{a,1}^2)\sigma_{a,0}^2/\sigma_{a,1}^2 \end{bmatrix}\right).$$

The fraction of the variation in $\{c_t\}$ due to future fundamentals is

$$\frac{\phi_1^2 \kappa^2}{\phi_0^2 + \phi_1^2 \kappa}. \tag{11}$$

For any positive values of $\sigma_{a,0}^2$ and $\sigma_{a,1}^2$, this fraction converges to zero as $\phi_1/\phi_0 \to 0$. As the endogenous variable places relatively less weight on expectations of future fundamentals compared to current fundamentals (i.e. as it becomes less forward-looking), the contribution of future fundamental shocks becomes smaller.

To find the contribution of news shocks, substitute $x_t = a_{0,t} + a_{1,t-1}$ and $E_t[x_{t+1}] = a_{1,t}$ into the expression for c_t in (10) to get $c_t = \phi_0 a_{0,t} + \phi_0 a_{1,t-1} + \phi_1 a_{1,t}$. Therefore the fraction of variation in $\{c_t\}$ due to news shocks is

$$\frac{(\phi_0^2 + \phi_1^2)\sigma_{a,1}^2}{\phi_0^2(\sigma_{a,0}^2 + \sigma_{a,1}^2) + \phi_1^2 \sigma_{a,1}^2}.$$

Comparing this expression with the expression in Equation (11), we can see that as ϕ_1/ϕ_0 approaches zero the contribution of future fundamentals approaches zero, while the contribution of news shocks approaches $\sigma_{a,1}^2/(\sigma_{a,0}^2 + \sigma_{a,1}^2)$. But this fraction can be driven arbitrarily close to one as $\sigma_{a,1}^2/\sigma_{a,0}^2$ approaches infinity. Therefore, the variation attributed to news shocks can be completely disconnected from the variation attributed to future fundamental shocks. They could be important mostly because *past* news shocks affect current fundamentals. Whether or not the importance of news shocks translates into the importance of future fundamental shocks therefore hinges crucially on the structure of one's economic model.

4 Structural VAR Analysis

A common view is that structural VAR analysis can be applied to models with news shocks, but not to models with noise shocks. The main reason is that news shocks can be expressed as a function of current and past observables, but noise shocks cannot. Noise representations are "non-invertible." By contrast, the reduced-form shocks from a VAR are a function of current and past observables. So, the thinking goes, it can't be appropriate to use an invertible VAR to say something about non-invertible noise shocks.

In this section, however, we argue that structural VAR analysis is equally applicable to models with news or noise shocks. We agree that noise representations are not invertible, but we disagree that invertibility is a necessary condition for using structural VARs. We prove that both news and noise shocks can be recovered from reduced-form VAR shocks up to one orthogonal transformation.

4.1 What is structural VAR analysis?

Structural VAR analysis has two steps: a VAR step and a structural step. The VAR step constructs a "reduced-form" representation of the observables with residuals that come from a projection of observables on their past history. The structural step uniquely determines a collection of "structural" shocks from the reduced-form residuals using one orthogonal transformation.

To be more precise, we can define a reduced-form representation of fundamentals and beliefs in the following way.[16]

Definition 4. In a "reduced-form" representation of fundamentals and beliefs,

$$\hat{x}_{i,t} = \tilde{x}_{i,t-1} + \tilde{\epsilon}_{i,t} \quad \text{for all } i, t \in \mathbb{Z},$$

where $\tilde{x}_{i,t-1} \in \mathcal{H}_{t-1}(\hat{x})$ and $\tilde{\epsilon}_{i,t} \perp \mathcal{H}_{t-1}(\hat{x})$.

The VAR step entails constructing this representation; that is, getting the reduced-form shocks $\{\tilde{\epsilon}_{i,t}\}$ from the observables $\{\hat{x}_{i,t}\}$. Notice that because these shocks are defined by an orthogonal projection, they are always unique. This representation is also called a VAR representation because we can think of it as a regression of the (potentially infinite-dimensional) vector $\hat{x}_t = (\ldots, \hat{x}_{-1,t}, \hat{x}_{0,t}, \hat{x}_{1,t}, \ldots)'$ on the space spanned by its own past values.

The second step entails recovering the collection of structural shocks from these reduced-form shocks using an orthogonal transformation. It is only possible to do this if the space spanned by the structural shocks is uniquely determined by the reduced-form shocks. The main theoretical question, then, is: do the reduced-form shocks uniquely determine the space spanned by the underlying structural shocks in both news and noise representations?

As we will see below, the answer is yes. But before presenting that result, we briefly discuss the concept of invertibility. We discuss the reasons why many researchers believe that structural VAR analysis is not applicable to models with noise shocks, and why we do not find those reasons convincing.

4.2 The "Problem" with Non-Invertibility

An important concept in discussions regarding the applicability of structural VAR analysis is that of invertibility. Invertibility has to do with whether the underlying shocks in a representation can be expressed as a function of current and past observables.

Definition 5. A collection of stochastic processes $\{y_{i,t}\}$, $i \in \mathcal{I}_y \subseteq \mathbb{Z}_+$ is "invertible" with respect to the collection of shocks $\{\epsilon_{i,t}\}$, $i \in \mathcal{I}_\epsilon \subseteq \mathbb{Z}_+$, if

$$\mathcal{H}_t(\epsilon) = \mathcal{H}_t(y) \quad \text{for all } t \in \mathbb{Z}.$$

Based on this definition, we refer to a representation of the collection $\{y_{i,t}\}$ as an "invertible representation" if $\{y_{i,t}\}$ is invertible with respect to all the underlying shocks in that repre-

[16] We continue to use Hilbert spaces here to allow for situations in which it might not be possible to summarize fundamentals and beliefs only by a finite-dimensional vector. We discuss the finite-dimensional case in Section (4.4).

sentation.[17] Also, note that we reserve the term "shock" to refer only to a process that is uncorrelated over time.

From Definition (4), it is easy to see that reduced-form representations are always invertible. Each reduced-form shocks is, by construction, a function of current and past observables. What about news and noise representations? The next result characterizes both representations in terms of invertibility.

Proposition 8. *News representations are invertible; noise representations are not.*

This result is a generalization of the one that Blanchard et al. (2013) prove in the context of a simple model of consumption determination, and the basic intuition is the same: if noise representations were invertible, then people would be able to distinguish the informative parts of their signals from the noise. By rationality, noise shocks could never affect people's beliefs. But then it would not be possible to recover those shocks from the current and past history of people's beliefs.

For many researchers, Proposition (8) settles the question of whether structural VAR analysis can be applied to models with noise shocks. From a strictly logical point of view, however, this proposition alone is not enough to rule out the possibility of applying structural VAR analysis to models with noise shocks. For it to do so, we must conjoin it with the premise

(P) Structural VAR analysis can only be applied to invertible representations.

The question is, should we accept (P)? It is difficult to find any explicit justification of this premise in the literature. In most studies that we are aware of, the authors simply take it for granted. For example, Fernández-Villaverde et al. (2007) frame their discussion around the question of whether a general linear economic model cast in state-space form is invertible, and present a sufficient condition for this to be true. Their discussion at the beginning of Section I.B appears to treat the questions of invertibility and the applicability of structural VAR analysis as equivalent.

We suspect that the reason this premise is often taken as self-evident is because in a non-invertible representation, the shocks must depend at least partly on *future* values of observables. But since the future is surely unknown to us, so the argument goes, it cannot be feasible to recover any non-invertible shocks using only current and past observations. In his discussion of non-invertibility, Hamilton (1994) expresses this view:

[17] What we call invertibility is sometimes called "fundamentalness" (cf. Rozanov, 1967, ch. 2). Unfortunately that terminology conflicts with our use of "fundamental" to describe the exogenous driving processes in an economic model, which is why we use "invertibility" instead.

> "To find the value of ϵ for date t associated with the invertible representation...we need to know current and past values of y. By contrast, to find the value of ϵ for date t associated with the non-invertible representation, we need to use all the future values of y! If the intention is to calculate the current value of ϵ_t using real-world data, it will be feasible only to work with the invertible representation." (p.67)

This sort of reasoning, however, seems to involve a confusion about what exactly "date t" refers to. On the one hand, there is the date at which a particular random variable is realized. On the other hand, there is the date at which a hypothetical econometrician performs his analysis, given a sample of past realizations. Even though the econometrician may not be able to observe the value of any realization that will occur after the date at which he performs his analysis, he does observe realizations that occur after each particular date within his sample (except perhaps the final one). As a result, he can be said to observe the future, *relative* to certain dates in his sample. Indeed, for every pair of observations in his sample, one is past and one is future relative to the other.

To clarify this point further, it can be helpful to present counterexamples in which a structural model is not invertible, but structural VAR analysis can nevertheless be applied. The following univariate example illustrates one such situation.

Example 1. Suppose that theory delivers the following economic model relating an endogenous process $\{y_t\}$ to a sequence of structural shocks $\{\epsilon_t\}$:

$$y_t = \alpha y_{t+1} + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), \tag{12}$$

with $0 < \alpha < 1$. This representation is not invertible, because $\epsilon_t = y_t - \alpha y_{t+1}$ cannot be expressed as a function of current and past values of $\{y_t\}$.

Now suppose that an econometrician is interested in using structural VAR analysis to say something about the structural shocks given observations of $\{y_t\}$. First, he constructs the reduced-form autoregressive representation

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + \tilde{\epsilon}_t, \quad \tilde{\epsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \tilde{\sigma}^2), \tag{13}$$

where the polynomial $\rho(z) \equiv 1 - \rho_1 z - \cdots - \rho_p z^p$ doesn't have any zeros inside the unit circle. Under the null hypothesis that the economic model is the true data generating process, the econometrician would find $\rho_1 = \alpha$, $\rho_j = 0$ for all $j > 1$, and $\tilde{\sigma}^2 = \sigma^2$. But because this is the reduced-form "VAR" step, he is agnostic about whether or not those relations are satisfied.

To identify the structural shocks, the econometrician needs some additional restrictions;

this is the "structural" step. Instead of imposing *all* the restrictions implied by the economic model, he only imposes a weaker set of restrictions that would nevertheless be *sufficient* to identify the structural shocks under the null. This weaker set takes the form of orthogonality conditions (zero restrictions). Specifically, he uses

$$E[y_t \epsilon_{t-1}] = 0 \quad \text{for all } j > 0$$

This says that current value of the endogenous process is not a function of past structural shocks. Because the economic model is purely forward looking, these conditions are sufficient to identify the space spanned by the structural shocks under the null. Since that space is one-dimensional, the choice of an orthogonal transformation amounts to a scale normalization. Suppose he selects the (correct) scale normalization $E[y_t \epsilon_t]/E[\epsilon_t^2] = 1$.

All that remains is to impose the structural restrictions on the reduced-form representation. In this case, the unique representation that is observationally equivalent to (13) and satisfies the structural restrictions is:

$$y_t = \rho_1 y_{t+1} + \rho_2 y_{t+2} + \cdots + \rho_p y_{t+p} + \hat{\epsilon}_t, \quad \hat{\epsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \tilde{\sigma}^2).$$

Notice that under the null, $\hat{\epsilon}_t = \epsilon_t$, so the econometrician would successfully recover the structural shocks using this structural VAR procedure. △

This counterexample is enough for us to conclude that (P) is not a valid premise. However, that conclusion only opens the door to the *possibility* that we may be able to use structural VAR analysis in models with noise shocks. In order to guarantee that structural VAR analysis is applicable in those cases, we will need more than just a counterexample. We turn to this issue in the next subsection.

4.3 Recovering News and Noise Shocks

Here we present the main theoretical result of this section, which is that structural VAR analysis is equally applicable to models with news or noise shocks. This is because in both cases, the underlying shocks are uniquely defined by the reduced-form shocks, up to an orthogonal transformation.

Proposition 9. *The shocks $\{\tilde{\epsilon}_{i,t}\}$ from the reduced-form representation in Definition* (4) *uniquely determine the space spanned by the underlying shocks at each date in any news or noise representation.*

Given the reduced-form shocks, all that is required to infer the values of news or noise shocks, or impulse responses to them, is one orthogonal transformation. In the case of a news representation, this orthogonal transformation is already built into Definition (1). The shocks are orthogonalized so that each news shock is orthogonal to all the rest, and so that the news shock associated with horizon $i \in \mathcal{I}$ has a unit impact on the forecast $\hat{x}_{i,t}$, but zero impact response on forecasts $\hat{x}_{j,t}$ for $j < i$. Under this orthogonalization, Proposition (9) can be strengthened to say that the reduced-form shocks uniquely determine the underlying shocks in the news representation.

In the case of a noise representation, there isn't any specific orthogonalization scheme built into Definition (2). So we have the flexibility to choose one to aid interpretation on a case by case basis. But to pick out one natural scheme that might be appropriate across a wide variety of applications, we can impose a recursive ordering analogous to the one used in the news representation. Select an orthogonal basis for the space spanned by noise shocks, and normalize the shocks so that noise shock i has a unit impact on forecast $\hat{x}_{i,t}$ but zero impact response on forecasts $\hat{x}_{j,t}$ for $j < i$. The interpretation of noise shock i under this scheme is that it generates a one-for-one increase at date t in the expectation of the fundamental realization at date $t+i$ which is unrelated to fundamentals shocks at any date, and which is unrelated to noise shocks that affect expectations of fundamentals between dates t and $t+i$. We can formalize this noise shock orthogonalization scheme as follows.

Assumption 1. *In any noise representation of fundamentals and beliefs, the following conditions are satisfied:*

(a) Orthogonal basis: $E[\epsilon^v_{i,t}\epsilon^v_{j,t}] = 0$ *for all* $i \neq j \in \mathcal{I}$,

(b) Scale normalization: $E[\hat{x}_{i,t}\epsilon^v_{i,t}]/E[(\epsilon^v_{i,t})^2] = 1$ *for all* $i \in \mathcal{I}$, *and*

(c) Recursive ordering: $E[\hat{x}_{j,t}\epsilon^v_{i,t}] = 0$ *for all* $j < i \in \mathcal{I}$.

Under this assumption, it follows that all the shocks in any noise representation can be uniquely recovered (along with their associated impulse response functions) from the collection of reduced-form shocks.

4.4 The Bivariate Case

To make matters concrete, in this subsection we demonstrate how structural VAR analysis can be applied in the bivariate context to recover news and noise shocks. In doing so, we also translate the preceding discussion into the language of lag polynomials, zero restrictions, and generating functions. This serves to highlight the parallels between the structural restrictions implied by news and noise models.

Suppose that fundamentals and beliefs about future fundamentals are completely summarized by the two-dimensional vector $d_t \equiv (x_t, \hat{x}_t)'$. The process $\{d_t\}$ is stationary and regular, and has a rational covariance generating function, $g_d(z)$ for $z \in \mathbb{C}$.

News and noise representations take a particularly simple form in this bivariate setting. The news representation expresses d_t in terms of the news and surprise shocks $\{\epsilon_t^a\}$ through a moving average of the form:

$$d_t = B(L)\epsilon_t^a, \quad \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_a), \tag{14}$$

where (i) $B(z)$ has no negative powers of z and $|B(z)|$ has no zeros inside the unit disk, (ii) $B(0)$ is upper triangular with unit diagonal elements, (iii) Σ_a is a diagonal matrix, and (iv) $g_d(z) = B(z)\Sigma_a B(z^{-1})'$.

Restriction (i) ensures that the representation is invertible, as Proposition (8) indicates must always be the case for a news representation. Restriction (ii) says that the surprise shock has a unit impact effect on fundamentals and the news shock has a unit impact effect on expectations, but that the news shock cannot affect fundamentals on impact. The impact effect of a surprise shock on expectations is not restricted. Restriction (iii) says that the two shocks are independent, and restriction (iv) merely guarantees that this is in fact a representation of $\{d_t\}$. We can rewrite (14) with these restrictions as follows:

$$\begin{bmatrix} x_t \\ \hat{x}_t \end{bmatrix} = \cdots + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}}_{B_{-1}} \begin{bmatrix} \epsilon_{0,t+1}^a \\ \epsilon_{1,t+1}^a \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix}}_{B_0} \begin{bmatrix} \epsilon_{0,t}^a \\ \epsilon_{1,t}^a \end{bmatrix} + \underbrace{\begin{bmatrix} * & * \\ * & * \end{bmatrix}}_{B_1} \begin{bmatrix} \epsilon_{0,t-1}^a \\ \epsilon_{1,t-1}^a \end{bmatrix} + \cdots,$$

and $\Sigma_a = \text{diag}(\sigma_{a,0}^2, \sigma_{a,1}^2)$.

Turning to the noise representation, the first issue that must be addressed is the fact that noise representations are not unique according to Definition (2). However, we can adopt the orthogonalization scheme from Assumption (1) to pin down one particular noise representation. Because the example is bivariate, the space spanned by noise shocks is only one dimensional. Therefore parts (a) and (c) of Assumption (1) are unnecessary, and adopting this assumption amounts to the scale normalization that a unit noise shock has a unit impact effect on expectations.

The noise representation of $\{d_t\}$ consistent with Assumption (1) expresses d_t in terms of the fundamental and noise shocks $\{\epsilon_t^s\}$ through a moving average:

$$d_t = C(L)\epsilon_t^s, \quad \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_s), \tag{15}$$

where (i) the diagonal elements of $C(z)$ have no negative powers of z and $|C(z)|$ has no zeros inside the unit disk, (ii) $C(0)$ has unit diagonal elements and $C(z)$ is lower triangular for all z, (iii) Σ_s is a diagonal matrix, and (iv) $g_d(z) = C(z)\Sigma_s C(z^{-1})'$.

Restriction (i) does not require this representation to be invertible, and expectations may depend on future values of the fundamental shocks. Restriction (ii) says that the fundamental shock has a unit impact effect on fundamentals, and the noise shock has a unit impact effect on expectations, but that the noise shock cannot affect the fundamental process at any lead or lag. Restrictions (iii) and (iv) are exactly the same as in the news representation. Rewriting (15) with these restrictions:

$$\begin{bmatrix} x_t \\ \hat{x}_t \end{bmatrix} = \cdots + \underbrace{\begin{bmatrix} 0 & 0 \\ * & 0 \end{bmatrix}}_{C_{-1}} \begin{bmatrix} \epsilon^x_{t+1} \\ \epsilon^v_{t+1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix}}_{C_0} \begin{bmatrix} \epsilon^x_t \\ \epsilon^v_t \end{bmatrix} + \underbrace{\begin{bmatrix} * & 0 \\ * & * \end{bmatrix}}_{C_1} \begin{bmatrix} \epsilon^x_{t-1} \\ \epsilon^v_{t-1} \end{bmatrix} + \cdots,$$

and $\Sigma_s = \text{diag}(\sigma_x^2, \sigma_v^2)$.

Now, suppose that we have an estimate of the generating function $g_d(z)$. Finding the news representation in (14) and the noise representation in (15) can be reduced to the problem of factoring $\{d_t\}$ in different ways. This is most easily seen through Restriction (iv) from both representations:

$$g_d(z) = \underbrace{B(z)\Sigma_a B(z^{-1})'}_{\text{news}} = \underbrace{C(z)\Sigma_s C(z^{-1})'}_{\text{noise}}.$$

From Theorem (1) we already know that it is always possible to achieve either factorization, and the proof of that theorem provides an algorithm for doing so. However, we can also provide an alternative algorithm that works directly on $g_d(z)$, and which can help to further clarify our earlier results.

First consider how to find the news representation. It is well known that $g_d(z)$ can always be uniquely factored in the form

$$g_d(z) = H(z)VH(z^{-1})^{-1}, \tag{16}$$

where (i) $H(z)$ has no negative powers of z and no zeros inside the unit disk, (ii) $B(0)$ is the identity matrix, and (iii) V is a symmetric matrix. This is the canonical factorization of $g_d(z)$, which corresponds to the Wold representation of $\{d_t\}$. An algorithm for finding $H(z)$ and V can found in Rozanov (1967, pp. 44-47).

The factors of the news representation, $B(z)$ and Σ_a, are closely related to $H(z)$ and V. Let $MDM' = V$ denote the (unique) LDL decomposition of V, and set $\Sigma_a = D$ and $B(z) = H(z)M$. It is easy to verify that all the requisite properties of $B(z)$ and Σ_a are satisfied, and therefore that we have found a way to go from $g_d(z)$ to the news representation.

Next, consider how to find the noise representation. Writing $g_d(z) = C(z)\Sigma_a C(z^{-1})'$ out more explicitly,

$$\begin{bmatrix} g_x(z) & g_{x\hat{x}}(z) \\ g_{\hat{x}x}(z) & g_{\hat{x}}(z) \end{bmatrix} = \begin{bmatrix} \sigma_x^2 C_{11}(z)^2 & \sigma_x^2 C_{11}(z) C_{21}(z^{-1}) \\ \sigma_x^2 C_{11}(z) C_{21}(z) & \sigma_v^2 C_{22}(z)^2 + \sigma_x^2 C_{21}(z) C_{21}(z^{-1}) \end{bmatrix}. \tag{17}$$

By restrictions (i) and (ii), the upper-left equation in this system says that σ_x^2 and $C_{11}(z)$ are none other than the canonical factors of $g_x(z)$. This means that the first equation in the system (14) is the univariate Wold representation of $\{x_t\}$, and $\{\epsilon_t^x\}$ are Wold innovations. Therefore σ_x^2 and $C_{11}(z)$ can be obtained using the same algorithm we used to find the multivariate Wold representation above.

The lower-left equation in (17) determines $C_{21}(z)$ as a function of $g_{\hat{x}x}(z)$ and $C_{11}(z)$ and σ_x^2, which have already been determined by the upper-left equation. The lower-right equation in (17) implies that $\sigma_v^2 C_{22}(z)^2 = g_{\hat{x}}(z) - \sigma_x^2 C_{21}(z)^2$. Together with restrictions (i) and (ii), this means that σ_v^2 and $C_{22}(z)$ are uniquely defined as the canonical factors of $g_{\hat{x}}(z) - \sigma_x^2 C_{21}(z)^2$. Therefore, we have uniquely determined the factors $C(z)$ and Σ of the noise representation.

So far, we have shown how to find news and noise representations given an estimate of the generating function $g_d(z)$. In the context of structural VAR analysis, this estimate comes from fitting a VAR model to the data:

$$A(L)d_t = \tilde{\epsilon}_t, \quad \tilde{\epsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma), \tag{18}$$

where $A(z)$ is an finite-order lag polynomial of degree p. The estimate of $g_d(z)$ implied by this VAR is $g_d(z) = A(z)^{-1}\Sigma[A(z^{-1})^{-1}]'$.

Using finite-order VARMA models to characterize the time-series properties of stochastic processes is in keeping with the tradition of Box and Jenkins (1970). One reason that structural VAR analysis has focused exclusively on autoregressive models of the type in (18) is because models of that type can be quickly and efficiently estimated using ordinary least squares. Another is that incorporating moving average terms can raise problems regarding the uniqueness of reduced-form parameters. However, it should be clear from the discussion in this subsection that $g_d(z)$ could in principle come from something other than a VAR.

To summarize, we have demonstrated how structural VAR analysis can be used to recover news and noise shocks in the bivariate case. Estimation of the VAR model implies an estimate of the generating function of the data. This generating function can then be factored to find either the news or noise representation.

4.5 A Monte Carlo Study

To build confidence in our claim that structural VAR analysis is applicable to models with noise shocks, we apply the approach described in the previous subsection to a simple bivariate model of consumption determination. The model comes from Section I of Blanchard et al. (2013), and describes the evolution of aggregate consumption $\{c_t\}$ in terms of aggregate productivity $\{x_t\}$. This example is illustrative because it is analytically convenient, and because it is the example used by Blanchard et al. (2013) to argue that it is *not* possible to use structural VAR analysis on models with noise shocks.

The model specifies that at each date, consumption is equal to the long-run conditional expectation of future productivity, $c_t = \lim_{j\to\infty} E_t[x_{t+j}]$, where the expectation is conditional on the current and past history of productivity and signals about future productivity, $\{x_\tau, s_\tau\}$ for all $\tau \leq t$. The signal process $\{s_t\}$ is related to future fundamentals according to the system of equations in (8).

We have already derived an observationally equivalent noise representation for this information structure in Proposition (7). In terms of that representation, consumption can be expressed as a function of current and expected future values of the process $\{m_t\}$:

$$c_t = \frac{1}{(1-\rho)^2} E_t[m_{t+2} - 2\rho m_{t+1} + \rho^2 m_t].$$

This equation and the system of equations in (8) together make up the true data generating process linking consumption to noise shocks. From this representation it is possible to compute various objects of interest, such as the impulse response function of consumption with respect to noise shocks, or the importance of noise shocks for explaining consumption over business cycle frequencies.

Our Monte Carlo exercise entails simulating data on consumption and productivity from this model, and placing ourselves in the shoes of an econometrician who has no knowledge of the true data generating process. He receives a finite sample of realizations, and is charged with estimating the importance of noise shocks and the effects of a noise shock on consumption from that sample. In practice, we simulate $N = 1000$ samples of $T = 275$ observations of consumption and productivity from this model. The reduced-form model is an unrestricted VAR, and the structural parameters are $\rho = 0.891$, $\sigma_\lambda = 0.07303$, and $\sigma_v = 0.89$, which are the same values chosen by Blanchard et al. (2013). The lag length of the VAR is chosen to minimize the information criterion proposed by Hannan and Quinn (1979).

The left panel of Figure (2) plots the true impulse response of consumption to a noise shock that increases consumption by one unit on impact, together with 95% bands constructed from the point estimates across the $N = 1000$ different samples. The true response

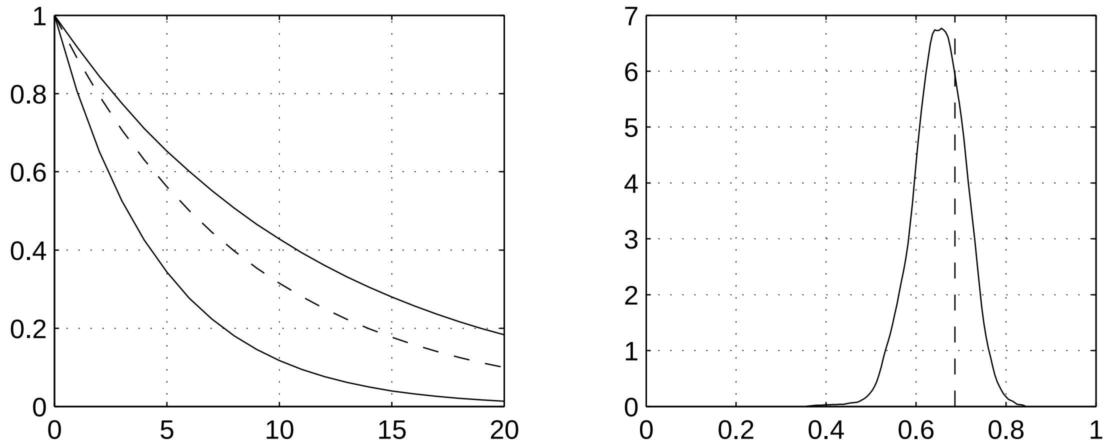

Figure 2: Strutural VAR analysis of data simulated from a model with noise shocks. *Left*: the dashed line is the true impulse response of consumption to a unit noise shock, while solid lines are 95% bands from the distribution of point estimates from each of $N = 1000$ samples of length $T = 275$. *Right*: the dashed line is the true contribution of noise shocks over business-cycle frequencies (6 to 32 quarters), and solid line is the distribution of point estimates over all simulated samples.

of consumption is one of geometric decay; initially consumption increases due to positive expectations about future productivity, but over time those effects die out as people come to realize that their expectations had only responded to noise. In the long run, the effect of noise shocks on consumption converges to zero.

The figure indicates that structural VAR analysis does a good job capturing the response of consumption to a noise shock, even for samples of $T = 275$ observations. Not surprisingly, increasing the sample size increases the accuracy of our estimates.

Perhaps one puzzling aspect of this result is that it is apparently possible for an econometrician to identify the effects of a shock that has non-flat effects on consumption. Blanchard et al. (2013) explain that an econometrician with access to the same information consumers or less, cannot identify any shock with non-flat effects on consumption. This is because consumption is a random walk in this model, conditional on consumers' information. We agree with this result.

However, the crucial observation is that econometricians always have *more* information than consumers. In order for an econometrician to perform any sort of analysis on a sample of data, that data must have been realized at some date in the past. As a result, relative to consumers at each date in his sample (except perhaps the very last date, in the case of a very timely sample), he has access to more information about both consumption and productivity. It is by using this additional information that the econometrician can successfully identify the effects of noise shocks.

The right panel of Figure (2) plots the share of the variance in consumption explained by noise shocks over business cycle frequencies (6 to 32 quarters). The vertical dashed line is the true noise share (0.69), while the solid line is the histogram of point estimates from each of the $N = 1000$ different samples. Again, the structural VAR procedure evidently delivers fairly reliable estimates of the importance of noise shocks. Based on the distribution of point estimates, it appears that the estimated importance of noise shocks does exhibit some slight downward bias in finite samples. As we will see, however, a slight downward bias in this estimate would only strengthen the conclusions we reach in the next section.

4.6 Noise Shocks in U.S. Data

In this subsection, we use structural VAR analysis to estimate the dynamic response of consumption to noise shocks and the importance of noise shocks in U.S. data. We measure consumption by the natural logarithm of real per-capita personal consumption expenditure (NIPA table 1.1.6, line 2, divided by BLS seires LNU00000000Q) and productivity by the natural logarithm of utilization adjusted total factor productivity (Federal Reserve Bank of San Francisco). Our sample is 1948:Q1 to 2016:Q4, which gives $T = 276$ observations.

Before discussing our results, a cautionary remark is in order regarding the interpretation of noise shocks in real-world data. In the theoretical model of the previous section, productivity is the only fundamental process, and people have rational expectations. As a result, the only reason that consumption can possibly move without some corresponding movement in current, past, or future productivity is because of rational errors induced by noisy signals. In the data, it is plausible that consumption is driven by fundamentals other than productivity, by sunspots, or even by non-rational fluctuations in people's beliefs. Therefore, noise shocks should be interpreted broadly in this section as composite shocks that capture all *non-productivity* fluctuations in consumption.

Keeping that interpretation in mind, we turn to Figure (3). The upper left panel plots the estimated impulse response of consumption to a noise shock that increases consumption by one unit on impact. The response is hump-shaped, increasing for six quarters after the shock, and then slowly decaying back toward zero. The effect of noise shocks is also highly persistent; even after 20 quarters the response is still statistically different from zero. To the extent that these shocks do represent rational mistakes due to imperfect signals, the high persistence means that it must take a while for people to figure out that a mistake was made.

The upper right panel of Figure (3) plots the share of the variance in consumption explained by noise shocks over business cycle frequencies (6 to 32 quarters). The vertical dashed line is our point estimate (0.86), while the solid line is the histogram of point estimates

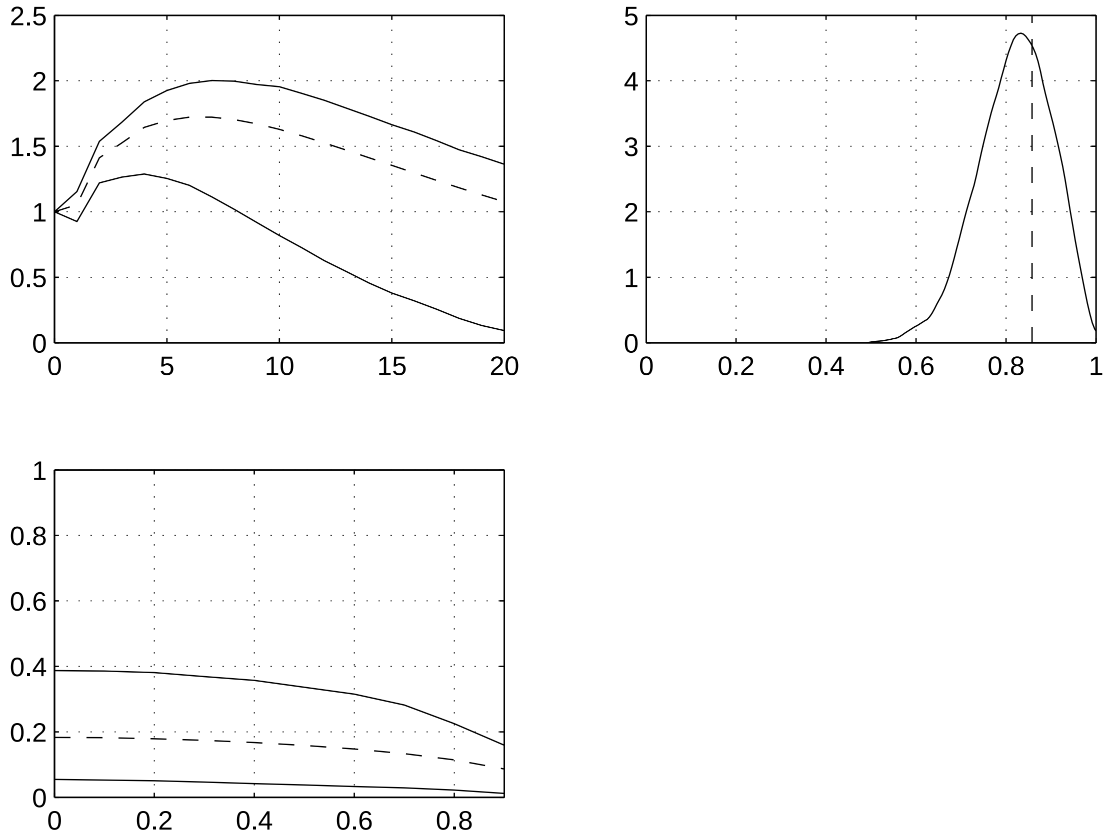

Figure 3: Structural VAR analysis of quarterly U.S. consumption and total factor productivity from 1948:Q1 to 2016:Q4. *Upper left*: response of consumption to unit noise shock. The dashed line is the point estimate, and the solid lines are 95% bootstrap confidence bands. *Upper right*: share of consumption variance due to noise shocks over business-cycle frequencies (6 to 32 quarters). The dashed line is the point estimate (0.86) and the solid line is the distribution of bootstrap estimates. *Lower left*: fraction of the fundamental share of consumption variance due to future fundamental shocks, as a function of the trend parameter $0 \leq \lambda < 1$. $\lambda = 0$ corresponds to a decomposition in (log) first differences, while $\lambda \to 1$ corresponds to a decomposition in (log) levels.

from each of the $N = 1000$ bootstrap samples. These results indicate that 86% of the business-cycle variation in consumption is due to noise shocks. An equivalent way of stating this is that productivity only explains 14% of the variation in consumption. This means that a large majority of consumption fluctuations are not due to productivity shocks.

Cochrane (1994) reaches a similar conclusion. Using structural VARs, he argues that the bulk of economic fluctuations is not due to productivity shocks (or a number of other shocks such as those due to monetary policy, oil prices, and credit). However, he does not control for the possibility that fluctuations might be due to *future* changes in productivity to which people respond in advance. We do control for this possibility. We find that while people's beliefs about future productivity may be moving around a lot, it appears either that those movements are mostly unrelated to subsequent changes in productivity, or that people's beliefs about future productivity do not matter very much for their current actions.

Furthermore, we can ask whether productivity matters mostly through current and past shocks, or through future shocks. The lower left panel of Figure (3) plots the faction of the productivity share due to future fundamental shocks, for a range of different de-trended versions of consumption. Future productivity shocks represent somewhere between 10% and 20% of the total contribution of productivity. As in the theoretical models from Section (3.5), this means that productivity affects consumption primarily through current and past shocks. Nevertheless, the dependence of consumption on future productivity shocks is not zero. There is some evidence that consumption contains information about future productivity beyond what can be known from the current and past history of productivity alone.

5 Conclusion

Models with news and noise are intimately related. In fact, as we have argued here, there is a precise sense in which they are identical. The missing link is the observation that they are really just two different ways of describing the joint dynamics of exogenous economic fundamentals and people's beliefs about them. This link is formalized by Theorem (1).

The observational equivalence between news and noise representations serves an important positive purpose. Namely, it provides a way to determine the importance of beliefs as an independent cause of fluctuations. A number of prominent studies have constructed models to understand how beliefs can drive fluctuations. However, because what these studies call "news" shocks actually mixes fluctuations due fundamentals and beliefs, none has fully isolated the contribution of beliefs that is independent of the contribution due to fundamentals.

In order to disentangle beliefs from fundamentals, we have shown how it is possible to derive a noise representation of the model, and then perform variance decompositions in terms of noise versus fundamental shocks. These decompositions are always unique by Proposition (4). We then apply our results to three quantitative models of the U.S. economy, from Schmitt-Grohé and Uribe (2012), Barsky and Sims (2012), and Blanchard et al. (2013). We show that these studies have all understated the importance of pure shocks to beliefs that is implied by their models.

We have further decomposed the contribution of fundamentals in a noise representation into the fraction that is due to future fundamental shocks, and that which is due to current and past fundamental shocks. All three of the quantitative models we consider imply that future fundamentals are much less important than current and past fundamentals. This result is consistent with the finding that news shocks play a large role because news shocks also capture fluctuations due to current and past fundamental shocks.

The observational equivalence of news and noise shocks also implies that structural VAR analysis is equally applicable to models with news or noise. Although we prove in Proposition (8) that noise models are non-invertible, invertibility is not a necessary condition for using structural VARs. An econometrician has more information than people in the economy have at the time they make their decisions. Therefore, he can use future observations to validate *ex post* whether those decisions can be justified by subsequent fundamental developments. In Proposition (9), we prove that the shocks in both news and noise representations can be recovered from reduced-form VAR shocks up to one orthogonal transformation.

In addition to showing that structural VAR analysis is applicable to noise models in principle, we also demonstrate how this can be done in practice. We use structural VARs to estimate the dynamic effects of noise shocks on consumption, along with their overall share in consumption fluctuations, in data simulated from a simple bivariate model of consumption determination with noise shocks, as well as in actual U.S. data on consumption and productivity. The simulation exercise allows us to verify that structural VARs do a good job recovering impulse responses, even in finite samples. In the application to actual data, we find that the large majority of fluctuations in consumption, 86%, are not due to productivity. In order to be consistent with this estimate and with an important role for shifting beliefs about productivity, it must be that fluctuations in beliefs about productivity are largely unrelated to aggregate productivity itself.

Our analysis of the importance of productivity shocks for explaining fluctuations in consumption invites a more thorough investigation into the sources of business cycle fluctuations. How important are other fundamentals, like monetary policy shocks, oil price shocks, credit shocks, or government spending shocks? What about other macroeconomic variables of interest like output, inflation, or unemployment? Of course, these questions are not new. However, the empirical procedure we develop in this paper can be helpful for determining the importance of a particular identifiable set of fundamental processes. This suggests the possibility of a new way forward on these old questions, which we leave for future work.

References

Anderson, B. D. and J. B. Moore (1979). *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall.

Angeletos, G.-M. and J. La'O (2013, March). Sentiments. *Econometrica 81*(2), 739–779.

Barsky, R. B., S. Basu, and K. Lee (2015, July). Whither News Shocks? In *NBER Macroeconomics Annual*, Volume 29, pp. 225–264. Chicago: University of Chicago Press.

Barsky, R. B. and E. R. Sims (2011). News shocks and business cycles. *Journal of Monetary Economics 58*(3), 273–289.

Barsky, R. B. and E. R. Sims (2012). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review 102*(4), 1343–1377.

Beaudry, P. and B. Lucke (2010, April). *Letting Different Views about Business Cycles Compete*, pp. 413–455. University of Chicago Press.

Beaudry, P., D. Nam, and J. Wang (2011, December). Do Mood Swings Drive Business Cycles and is It Rational? Working Paper 17651, National Bureau of Economic Research.

Beaudry, P. and F. Portier (2006). Stock prices, news, and economic fluctuations. *American Economic Review 96*(4), 1293–1307.

Beaudry, P. and F. Portier (2014, December). News-Driven Business Cycles: Insights and Challenges. *Journal of Economic Literature 52*(4), 993–1074.

Benhabib, J., P. Wang, and Y. Wen (2015). Sentiments and Aggregate Demand Fluctuations. *Econometrica 83*(2), 549–585.

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni (2013). News, noise, and fluctuations: An empirical exploration. *American Economic Review 103*(7), 3045–3070.

Born, B., A. Peter, and J. Pfeifer (2013, 12). Fiscal News and Macroeconomic Volatility. *Journal of Economic Dynamics and Control 37*(12), 2582–2601.

Box, G. and G. Jenkins (1970). *Stationary Random Processes*. San Francisco, CA: Holden-Day.

Christiano, L. J., C. Ilut, R. Motto, and M. Rostagno (2010). Monetary policy and stock market booms. *Proceedings of the Jackson Hole Economic Policy Symposium 34*(1), 85–145.

Christiano, L. J., R. Motto, and M. Rostagno (2014). Risk shocks. *The American Economic Review 104*(1), 27–65.

Cochrane, J. H. (1994, December). Shocks. *Carnegie-Rochester Conference Series on Public Policy 41*(1), 295–364.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007, June). Abcs (and ds) of understanding vars. *American Economic Review 97*(3), 1021–1026.

Forni, M., L. Gambetti, M. Lippi, and L. Sala (2014). Noisy News in Business Cycles. *Baffi Center Research Paper* (2014-159).

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J. D. (2017, January). Why you should never use the hodrick-prescott filter. Working paper, UC San Diego.

Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological) 41*(2), 190–195.

Jaimovich, N. and S. Rebelo (2009). Can news about the future drive the business cycle? *American Economic Review 99*(4), 1097–1118.

Jinnai, R. (2014). R&d shocks and news shocks. *Journal of Money, Credit and Banking 46*(7), 1457–1478.

Kurmann, A. and C. Otrok (2013). News shocks and the slope of the term structure of interest rates. *The American Economic Review 103*(6), 2612–2632.

Leeper, E. M. and T. B. Walker (2009). Information Flows and News Driven Business Cycles. Manuscript.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013). Fiscal Foresight and Information Flows. *Econometrica 81*(3), 1115–1145.

Lorenzoni, G. (2009). A theory of demand shocks. *American Economic Review 99*(5), 2050–84.

Lorenzoni, G. (2011). News and aggregate demand shocks. *Annual Review of Economics 3*(1), 537–557.

Pigou, A. C. (1927). *Industrial Fluctuations*. London: Macmillan.

Rozanov, Y. A. (1967). *Stationary Random Processes*. San Francisco, CA: Holden-Day.

Sargent, T. J. (1987). *Macroeconomic Theory: Second Edition*. Bingley: Emerald Group.

Schmitt-Grohé, S. and M. Uribe (2012). What's news in business cycles. *Econometrica 80*(6), 2733–2764.

Sims, E. (2016). What's news in news? a cautionary note on using a variance decomposition to assess the quantitative importance of news shocks. *Journal of Economic Dynamics and Control 73*(1), 41–60.

Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*. Minneapolis, MN: University of Minnesota Press.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala: Almqvist and Wiksell.

Appendix

Proof of Proposition (1)**.** Let $\hat{x}_t \equiv E_t[x_{t+1}]$ denote people's expectations of the fundamental at date $t+1$ given their information at t. The observable processes are $\{x_t\}$ and $\{\hat{x}_t\}$. Expectations at horizons greater than one are spanned by these two processes.

The two representations are observationally equivalent if and only if the covariance generating function (c.g.f.) of the data vector $d_t \equiv (x_t, \hat{x}_t)'$ is the same under either representation. Let $g_d(z)$ denote the c.g.f. of d_t, where z is a number in the complex plane. Then we can equate the c.g.f.'s implied by each representation:

$$g_d(z) = \underbrace{\begin{bmatrix} \sigma_{a,0}^2 + \sigma_{a,1}^2 & \sigma_{a,1}^2 z \\ \sigma_{a,1}^2 z^{-1} & \sigma_{a,1}^2 \end{bmatrix}}_{\text{news}} = \underbrace{\begin{bmatrix} \sigma_x^2 & \left(\frac{\sigma_x^4}{\sigma_x^2+\sigma_v^2}\right) z \\ \left(\frac{\sigma_x^4}{\sigma_x^2+\sigma_v^2}\right) z^{-1} & \left(\frac{\sigma_x^4}{\sigma_x^2+\sigma_v^2}\right) \end{bmatrix}}_{\text{noise}}.$$

This equality holds if and only if the relations in Proposition (1) are satisfied. □

Proof of Theorem (1)**.** To prove the first part, note that because $\mathcal{H}_{t-1}(\hat{x}) \subset \mathcal{H}_t(\hat{x})$ for all $t \in \mathbb{Z}$, it is possible to decompose $\mathcal{H}_t(\hat{x})$ into an orthogonal family of subspaces $\mathcal{H}_t(\hat{x}) = \bigoplus_{i=0}^{\infty} \mathcal{D}_{t-i}(\hat{x})$, where $\mathcal{D}_t(\hat{x}) \equiv \mathcal{H}_t(\hat{x}) \ominus \mathcal{H}_{t-1}(\hat{x})$ (cf. Rozanov, 1967, ch. 2). This means that $x_t \in \mathcal{H}_t(\hat{x})$ has a unique representation of the form

$$x_t = \sum_{i=0}^{\infty} w_{i,t-i}, \tag{19}$$

where the random variable $w_{i,t-i}$ represents the projection of x_t onto $\mathcal{D}_{t-i}(\hat{x})$ for any $i \in \mathbb{Z}_+$. By the orthogonality of the sequence of subspaces $\{\mathcal{D}_t(\hat{x})\}$, the process $\{w_{i,t}\}$ is uncorrelated over time for each $i \in \mathbb{Z}_+$.

While equation (19) looks almost like a news representation, it does not satisfy Definition (1) because it may be that $w_{i,t} \not\perp w_{j,t}$ for some $i \neq j$. Therefore, we use a version of the Gram-Schmidt orthogonalization procedure to transform these into an orthogonal sequence of shocks. Specifically, we define:

$$\epsilon_{0,t}^a = w_{0,t} \quad \epsilon_{i,t}^a = w_{i,t} - \sum_{j=0}^{i-1} \phi_{i,j} \epsilon_{j,t}^a \quad \text{for } i > 0,$$

where $\phi_{i,j} \equiv \langle w_{i,t}, \epsilon_{j,t}^a \rangle / \|\epsilon_{j,t}^a\|^2$ is a projection coefficient. Define the index set $\mathcal{I}$ to be the set of indices $i \in \mathbb{Z}_+$ such that $\|\epsilon_{i,t}^a\| > 0$. The collection of orthogonal shocks $\epsilon_{i,t}$ with $i \in \mathcal{I}$ is uniquely determined because the collection of input shocks $w_{i,t}$ with $i \in \mathbb{Z}_+$ is unique.

Substituting the orthogonalized shocks into equation (19), x_t can be uniquely rewritten as:

$$x_t = \sum_{i=0}^{\infty} \sum_{j \le i} \phi_{i,j} \epsilon^a_{j,t-i} = \sum_{j \in \mathcal{I}} \sum_{i=j}^{\infty} \phi_{i,j} \epsilon^a_{j,t-i} = \sum_{j \in \mathcal{I}} a_{j,t-j}.$$

The second equality rearranges the indexes on the double summation, and the third equality introduces the definition $a_{j,t-j} \equiv \sum_{i=j}^{\infty} \phi_{i,j} \epsilon^a_{j,t-i}$. The fact that the orthogonalized shocks are also uncorrelated over time implies that $a_{j,t} \perp a_{k,\tau}$ for all $j \neq k$ and $t, \tau \in \mathbb{Z}$. Therefore, this defines the unique news representation when people's date-t information set is $\mathcal{H}_t(a)$.

What remains is to prove that the expectations implied by this news representation are in fact equal to $\{\hat{x}_{i,t}\}$ for any $i \in \mathbb{Z}$. Under rational expectations, the i-step ahead expectation of x_t at date t under the original noise representation is equal to the orthogonal projection of x_{t+i} onto $\mathcal{H}_t(\hat{x})$: $\hat{x}_{j,t} = E[x_{t+j}|\mathcal{H}_t(\hat{x})]$. By the uniqueness of orthogonal projections,

$$w_{i,t} = \hat{x}_{i,t} - \hat{x}_{i+1,t-1},$$

where $w_{i,t}$ was defined in equation (19). Therefore, $\mathcal{H}_t(w) = \mathcal{H}_t(\hat{x})$. But then because $\mathcal{H}_t(a) = \mathcal{H}_t(w)$ by construction, it follows that $\mathcal{H}_t(a) = \mathcal{H}_t(\hat{x})$. So expectations are indeed the same under both representations, $\hat{x}_{i,t} = E[x_{t+i}|\mathcal{H}_t(\hat{x})] = E[x_{t+i}|\mathcal{H}_t(a)]$, which completes the proof of the first part of the theorem.

To prove the second part, we start from the (unique) news representation and define

$$s_{i,t} \equiv a_{i,t} \quad \text{for all } i \in \mathcal{I}.$$

Because $\mathcal{H}(x) \subset \mathcal{H}(a)$, there exist unique elements $m_{i,t} \in \mathcal{H}(x)$ and $v_{i,t} \in \mathcal{H}(s) \ominus \mathcal{H}(x)$ such that $s_{i,t} = m_{i,t} + v_{i,t}$. This defines a noise representation when people's date-t information set is $\mathcal{H}_t(s)$. What remains is to prove that the expectations implied by this noise representation are the same as the ones implied by the original news representation. Because $\mathcal{H}_t(s) = \mathcal{H}_t(a)$ by construction, and $\mathcal{H}_t(a) = \mathcal{H}_t(\hat{x})$ by the definition of a news representation, it follows that $\mathcal{H}_t(s) = \mathcal{H}_t(\hat{x})$ and therefore expectations are the same, $\hat{x}_{i,t} = E[x_{t+i}|\mathcal{H}_t(\hat{x})] = E[x_{t+i}|\mathcal{H}_t(s)]$. This completes the proof of the second part of the theorem. □

Appendix for Online Publication

A Proofs

Proof of Proposition (2)**.** Observational equivalence requires the probability distribution of the collection forecasts $\{\hat{x}_{i,t}\}$ for all $i, t \in \mathbb{Z}$ to be the same under both representations. The restriction that all random variables generating people's date-t information set are Gaussian, together with the rationality of expectations, implies that $\{\hat{x}_{i,t}\}$ fully characterizes the entire predictive distribution over realizations of the sequence $\{x_t\}$ at each date.

For each $t \in \mathbb{Z}$, let $\mathcal{F}_t(\hat{x})$ denote the smallest σ-algebra generated by $\{\hat{x}_{i,\tau}\}$ for $i, \tau \in \mathbb{Z}$ and $\tau \leq t$. That is, $\mathcal{F}_t(\hat{x})$ is generated by cylinder sets of the form

$$\mathcal{A}_t \equiv \{\omega \in \Omega : \hat{x}_{i_1,t_1} \in \mathcal{G}_1, \ldots, \hat{x}_{i_n,t_n} \in \mathcal{G}_n\},$$

where Ω denotes the space of elementary events, $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are arbitrary Borel sets in $\mathbb{R}$, the indices $t_1, \ldots, t_n$ assume values in the set $\{\tau \in \mathbb{Z} : \tau \leq t\}$, and the indices $i_1, \ldots, i_n$ assume values in $\mathbb{Z}$. Because each $\mathcal{F}_t(\hat{x})$ is unique, the sequence of σ-algebras $\{\mathcal{F}_t(\hat{x})\}$ is uniquely determined by the collection of forecasts $\{\hat{x}_{i,t}\}$. Therefore, the probability distribution over any stochastic process $\{c_t\}$, where for each $t \in \mathbb{Z}$ the random variable c_t is measurable with respect to $\mathcal{F}_t(\hat{x})$, is the same for any two observationally equivalent representations of fundamentals and beliefs. $\square$

Proof of Proposition (3)**.** As in the proof of Proposition (1), we can equate the c.g.f. of $d_t \equiv (x_t, \hat{x}_t)'$ that is implied by each representation:

$$g_d(z) = \underbrace{\begin{bmatrix} \sigma_\eta^2 + \sigma_\lambda^2 & \left(\frac{\sigma_\lambda^4}{\sigma_\lambda^2+\sigma_\xi^2}\right) z \\ \left(\frac{\sigma_\lambda^4}{\sigma_\lambda^2+\sigma_\xi^2}\right) z^{-1} & \frac{\sigma_\lambda^4}{\sigma_\lambda^2+\sigma_\xi^2} \end{bmatrix}}_{\text{system (5)}} = \underbrace{\begin{bmatrix} \sigma_x^2 & \left(\frac{\sigma_x^4}{\sigma_x^2+\sigma_v^2}\right) z \\ \left(\frac{\sigma_x^4}{\sigma_x^2+\sigma_v^2}\right) z^{-1} & \frac{\sigma_x^4}{\sigma_x^2+\sigma_v^2} \end{bmatrix}}_{\text{noise}}.$$

This equality holds if and only if the relations in Proposition (3) are satisfied. $\square$

Proof of Proposition (4)**.** Consider an arbitrary noise representation of fundamentals and beliefs and an arbitrary endogenous process $\{c_t\}$. Using the structure of signals in a noise representation, $\mathcal{H}(s) = \mathcal{H}(m) \oplus \mathcal{H}(v)$. Because $v_{i,t} \in \mathcal{H}(s) \ominus \mathcal{H}(x)$ for all $i \in \mathcal{I}$, the uniqueness of orthogonal decompositions implies that $\mathcal{H}(m) = \mathcal{H}(x)$. Therefore, $\mathcal{H}(s) = \mathcal{H}(x) \oplus \mathcal{H}(v)$. Furthermore, the definition of noise shocks implies that $\mathcal{H}(\epsilon^v) = \mathcal{H}(v)$, so

$$\mathcal{H}(s) = \mathcal{H}(x) \oplus \mathcal{H}(\epsilon^v). \tag{20}$$

By the endogeneity of $\{c_t\}$ and the rationality of expectations, $c_t \in \mathcal{H}(s)$ for all $t \in \mathbb{Z}$. Combining this with Equation (20), it follows that for each c_t, there exist two unique elements $a_t \in \mathcal{H}(x)$ and $b_t \in \mathcal{H}(\epsilon^v)$ such that

$$c_t = a_t + b_t.$$

To consider variance decompositions at different frequencies, let $f_y(\omega)$ denote the spectral density function of a stochastic process $\{y_t\}$. Then because $a_t \perp b_t$ for all $t \in \mathbb{Z}$, it follows that

$$f_c(\omega) = f_a(\omega) + f_b(\omega),$$

where the functions $f_a(\omega)$ and $f_b(\omega)$ are uniquely determined by the processes $\{a_t\}$ and $\{b_t\}$. These functions in turn uniquely determine the share of the variance of $\{c_t\}$ due to noise shocks in any frequency range $\underline{\omega} < \omega < \overline{\omega}$, which is equal to

$$\frac{\int_{\underline{\omega}}^{\overline{\omega}} f_b(\omega) d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} f_c(\omega) d\omega}.$$

The share due to fundamentals is equal to one minus this expression. □

Proof of Corollary (1)**.** Consider an arbitrary noise representation of fundamentals and beliefs, and an endogenous process $\{c_t\}$. By the rationality of expectations, people's best forecast of c_{t+h} as of date t is equal to

$$\hat{c}_{h,t} = E[c_{t+h}|\mathcal{H}_t(s)] = E[c_{t+h}|\mathcal{H}_t(\hat{x})].$$

Therefore, $\hat{c}_{h,t} \in \mathcal{H}_t(\hat{x})$. This means that the forecast error $w_t^h \equiv c_t - \hat{c}_{h,t-h}$ also satisfies $w_{h,t} \in \mathcal{H}_t(\hat{x})$. Therefore, $\{w_t^h\}$ is an endogenous process. By Proposition (4), the variance decomposition of this process in terms of noise and fundamentals is uniquely determined over any frequency range. Moreover, this result is true for any forecast horizon $h \in \mathbb{Z}$ because h was chosen arbitrarily. □

Lemma 1. *Any news representation in which each process $\{a_{i,t}\}$ is i.i.d. over time is observationally equivalent to a noise representation with $x_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_x^2)$ and*

$$s_{i,t} = x_{t+i} + v_{i,t}, \quad v_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{v,i}^2),$$

where $v_{i,t} \perp x_\tau$ and $v_{i,t} \perp v_{j,\tau}$ for any $i \neq j \in \mathcal{I}$ and $t, \tau \in \mathbb{Z}$, if and only if

$$\sigma_x^2 = \sum_{i \in \mathcal{I}} \sigma_{a,i}^2 \quad \text{and} \quad \sigma_{v,i}^2 = \frac{1}{\sigma_{a,i}^2} \left(\sum_{j<i} \sigma_{a,j}^2 \right) \left(\sum_{j \leq i} \sigma_{a,j}^2 \right) \quad \text{for all } i \in \mathcal{I}.$$

Proof of Lemma (1)**.** The proof of this result is a straightforward generalization of the proof of Proposition (1). In a news representation with i.i.d. news processes, the cross-c.g.f. of any two forecast processes $\{\hat{x}_{i,t}\}$ and $\{\hat{x}_{j,t}\}$ for $i, j \in \mathbb{Z}_+$ is equal to

$$g_{i,j}(z) = \sum_{k \in \mathcal{K}} \sigma_{a,k}^2 z^{j-i}, \tag{21}$$

where $\mathcal{K}$ is defined as the set of indices $k \in \mathcal{I}$ such that $k \geq |j-i|+i$. In a news representation of the type described in the proposition, the cross-c.g.f. of any two forecast processes $\{\hat{x}_{i,t}\}$ and $\{\hat{x}_{j,t}\}$ for $i, j \in \mathbb{Z}_+$ is equal to

$$\begin{aligned} g_{0,0}(z) &= \sigma_x^2 \\ g_{i,j}(z) &= \sigma_x^2 \left[1 + \frac{1/\sigma_x^2}{\sum_{k \in \mathcal{K}} 1/\sigma_{v,k}^2} \right]^{-1} z^{j-i} \quad \text{for } i, j > 0. \end{aligned} \tag{22}$$

Equating the c.g.f.'s in (21) with those in (22), and recursively solving for the parameters of the noise representation delivers the relations stated in the lemma. □

Proof of Proposition (5)**.** Define the composite shock

$$\epsilon_t^x \equiv \epsilon_{0,t}^a + \epsilon_{4,t-4}^a + \epsilon_{8,t-8}^a. \tag{23}$$

The process $\{\epsilon_t^x\}$ is i.i.d. because $\{\epsilon_{i,t}^a\}$ is i.i.d. for each $i \in \mathcal{I} \equiv \{0, 4, 8\}$. People's date-$t$ information set in representation (6) is $\mathcal{H}_t(\epsilon^a)$. But based on this information set, equation (23) defines a news representation for $\{\epsilon_t^x\}$ with i.i.d. news processes. Therefore, we can apply Lemma (1) to the composite shock process, which gives the relations stated in the corollary. Finally, note that $\hat{x}_{i,t} \in \mathcal{H}_t(\hat{\epsilon}^x)$ for all $t \in \mathbb{Z}$, which means that each $\{\hat{x}_{i,t}\}$ is endogenous with respect to $\{\epsilon_t^x\}$. Therefore, observational equivalence in terms of $\{\epsilon_t^x\}$ implies observational equivalence in terms of $\{x_t\}$. □

Proof of Proposition (6)**.** According to representation (7), the two signals observed by people in the economy are $s_{0,t} \equiv x_t$ and $s_{1,t} \equiv \lambda_t + \xi_t$. Because $\mathcal{H}(x) \subset \mathcal{H}(s)$, there exist two unique elements $m_t \in \mathcal{H}(x)$ and $v_t \perp \mathcal{H}(x)$ such that:

$$s_{1,t} = m_t + v_t \quad \text{for all } t \in \mathbb{Z}.$$

Using the finite-order ARMA restrictions in system (7), it follows that m_t is related to $\{x_t\}$ according to (cf. Whittle, 1983, ch. 5):

$$m_t = \alpha(L)x_t, \quad \alpha(z) \equiv \frac{\sigma_\lambda^2}{z[\sigma_\lambda^2 + \sigma_\eta^2(1-\rho z)(1-\rho z^{-1})]} \tag{24}$$

where L is the lag operator and $z \in \mathbb{C}$. We can factor this expression for $\alpha(z)$ as

$$\alpha(z) = \frac{\delta\sigma_\lambda^2}{\rho\sigma_\eta^2 z(1-\delta z)(1-\delta z^{-1})},$$

where $|\delta| < 1$ is equal to the expression stated in the proposition. By the definition of $\{v_t\}$,

$$g_v(z) = \frac{\sigma_\lambda^2\sigma_\eta^2}{\sigma_\lambda^2 + \sigma_\eta^2(1-\rho z)(1-\rho z^{-1})} + \sigma_\xi^2 = \sigma_\xi^2 \frac{\delta}{\beta} \frac{(1-\beta z)(1-\beta z^{-1})}{(1-\delta z)(1-\delta z^{-1})},$$

where $|\beta| < 1$ is equal to the expression stated in the proposition. This is the canonical form for $g_v(z)$ (cf. Whittle, 1983, ch. 2), which means that $\{v_t\}$ has an ARMA(1,1) representation. Because $\mathcal{H}_t(s)$ is unchanged from representation (7) for all $t \in \mathbb{Z}$, it follows that $\hat{x}_{i,t} \equiv E[x_{t+i}|\mathcal{H}_t(s)]$ is also unchanged for any $i \in \mathbb{Z}$. Therefore these two representations are observationally equivalent. □

Proof of Proposition (7)**.** A complication here is that both fundamentals and the signal of future fundamentals are difference-stationary, rather than stationary processes. As a result, they do not have finite second moments, which is a prerequisite for working in $\mathcal{L}^2$. However, this complication can be easily accommodated by defining the stationary processes $\{\tilde{x}_t\}$ and $\{\tilde{s}_t\}$ such that

$$(1-\lambda L)\tilde{x}_t = (1-L)x_t \quad \text{and} \quad (1-\lambda L)\tilde{s}_t = s_t,$$

where $0 \leq \lambda < 1$ and L is the lag operator. We can derive the noise representation in terms of $\{\tilde{x}_t\}$ and $\{\tilde{s}_t\}$ and then let λ tend to 1 from below (cf. Whittle, 1983, ch.8).

The two signals observed by people in the economy are therefore $\tilde{s}_{0,t} \equiv \tilde{x}_t$ and $\tilde{s}_{1,t} \equiv \tilde{s}_t$. Because $\mathcal{H}(\tilde{x}) \subset \mathcal{H}(\tilde{s})$, there exist two unique elements $\tilde{m}_t \in \mathcal{H}(\tilde{x})$ and $\tilde{v}_t \perp \mathcal{H}(\tilde{x})$ such that:

$$\tilde{s}_t = \tilde{m}_t + \tilde{v}_t \quad \text{for all } t \in \mathbb{Z}.$$

Using the finite-order ARMA restrictions in system (8), it follows that $\tilde{m}_t$ is related to $\{\tilde{x}_t\}$ according to

$$\tilde{m}_t = \alpha(L)\tilde{x}_t, \quad \alpha(z) \equiv \frac{\rho\sigma_\lambda^2/\sigma_\eta^2}{(1-\rho z)(1-\rho z^{-1})},$$

where $z \in \mathbb{C}$. By the definition of $\{\tilde{v}_t\}$,

$$g_{\tilde{v}}(z) = \frac{(1-z)(1-z^{-1})\left(\sigma_\xi^2[(1-\rho z)(1-\rho z^{-1})]^2 + \rho\sigma_\lambda^2\right)}{(1-\theta z)(1-\theta z^{-1})[(1-\rho z)(1-\rho z^{-1})]^2}.$$

The numerator can be factored as

$$\frac{\rho^2\sigma_\xi^2}{\delta\bar{\delta}}(1-\delta z)(1-\delta z^{-1})(1-\bar{\delta}z)(1-\bar{\delta}z^{-1}),$$

where $|\delta| < 1$ is equal to the expression stated in the proposition. This gives the canonical factorization of $g_{\tilde{v}}(z)$. Finally, we take limits as θ approaches 1 from below, and define variables without tildes as the limiting processes. Because $\mathcal{H}_t(s)$ is unchanged from representation (7) for all $t \in \mathbb{Z}$, it follows that $\hat{x}_{i,t} \equiv E[x_{t+i}|\mathcal{H}_t(s)]$ is also unchanged for any $i \in \mathbb{Z}$. Therefore these two representations are observationally equivalent. □

Proof of Proposition (8)**.** Begin with an arbitrary news representation. By definition of shocks, $\mathcal{H}_t(\epsilon^a) = \mathcal{H}_t(a)$. By the rationality of expectations, $\mathcal{H}_t(\hat{x}) = \mathcal{H}_t(a)$. Therefore, $\mathcal{H}_t(\epsilon^a) = \mathcal{H}_t(\hat{x})$, so this is an invertible representation.

Now consider an arbitrary noise representation. Suppose that it is invertible. Then $\epsilon_{i,t}^v \in \mathcal{H}_t(\hat{x})$ for any $i \in \mathcal{I} \subseteq \mathbb{Z}_+$ and $t \in \mathbb{Z}$. This implies that the noise shocks are contained in the information set of agents. But by rationality, if $\epsilon_{i,t}^v$ is contained in the information set of agents at date t, because it is uncorrelated with fundamentals, $\epsilon_{i,t}^v \notin \mathcal{H}_t(\hat{x})$. This is a contradiction. Therefore, the representation is not invertible. □

Proof of Proposition (9)**.** The definition of the reduced-form representation and the regularity of $\{\hat{x}_{i,t}\}$ imply that each $\hat{x}_{i,t}$ can be recovered from the current and past history of reduced-form shocks. Therefore, it is sufficient to prove that the space spanned by the underlying shocks in any news or noise representation is uniquely determined by the collection of observables $\{\hat{x}_{i,t}\}$.

Consider first the case of an arbitrary news representation of $\{\hat{x}_{i,t}\}$. By Theorem (1), this representation is unique, so *a fortiori* the space spanned by the underlying shocks in this representation is uniquely determined.

Now consider an arbitrary noise representation of $\{\hat{x}_{i,t}\}$. Because $\hat{x}_{i,t} \in \mathcal{H}(s) = \mathcal{H}(x) \oplus \mathcal{H}(v)$ (as we saw in the proof of Proposition (4)), it follows that $\hat{x}_{i,t}$ has a unique decomposition of the form

$$\hat{x}_{i,t} = \hat{m}_{i,t} + \hat{v}_{i,t},$$

where $\hat{m}_{i,t} \in \mathcal{H}(x)$ and $\hat{v}_{i,t} \perp \mathcal{H}(x)$ for any $i \in \mathbb{Z}$. To prove that $\mathcal{D}_t(v) \equiv \mathcal{H}_t(v) \ominus \mathcal{H}_{t-1}(v)$ is uniquely determined for all $t \in \mathbb{Z}$, it is sufficient to prove that $\mathcal{H}_t(v) = \mathcal{H}_t(\hat{v})$. The uniqueness of orthogonal decompositions then ensures that $\mathcal{D}_t(v) \equiv \mathcal{H}_t(v) \ominus \mathcal{H}_{t-1}(v) = \mathcal{H}_t(\hat{v}) \ominus \mathcal{H}_{t-1}(\hat{v}) \equiv \mathcal{D}_t(\hat{v})$.

Because $\mathcal{H}_t(s) = \mathcal{H}_t(\hat{x})$ by the definition of a noise representation, $s_{i,t}$ has a unique representation in terms of current and past expectations:

$$s_{i,t} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{i,j,k} \hat{x}_{k,t-j},$$

where $\psi_{i,j,k} \equiv \langle s_{i,t}, \hat{x}_{k,t-j} \rangle / \|\hat{x}_{k,t}\|^2$. Projecting both sides onto the subspace $\mathcal{H}(s) \ominus \mathcal{H}(x)$, and using the linearity of orthogonal projections, we can write

$$v_{i,t} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{i,j,k} \hat{v}_{k,t-j},$$

so $\mathcal{H}_t(v) \subseteq \mathcal{H}_t(\hat{v})$. Conversely, $\mathcal{H}_t(\hat{x}) = \mathcal{H}_t(s)$ implies that $\mathcal{H}_t(\hat{v}) \subseteq \mathcal{H}_t(v)$ because

$$\hat{v}_{i,t} = \sum_{j=0}^{\infty} \sum_{k \in \mathcal{I}} \theta_{i,j,k} v_{k,t-j},$$

where $\theta_{i,j,k} \equiv \langle \hat{x}_{i,t}, s_{k,t-j} \rangle / \|s_{k,t}\|^2$. Therefore, $\mathcal{H}_t(v) = \mathcal{H}_t(\hat{v})$. Finally, because $\hat{x}_{0,t} = x_t$ by $\mathcal{H}_t(x) \subset \mathcal{H}_t(s)$ and the rationality of expectations, it follows that the space spanned by fundamental shocks is uniquely determined as well. □

B Quantitative Models

The following subsections provide a sketch of each of the three quantitative models considered in this paper. For more details, we refer the reader to the original articles and their supplementary material.

B.1 Model of Schmitt-Grohé and Uribe (2012)

A representative household chooses consumption $\{C_t\}$, labor supply $\{h_t\}$, investment $\{I_t\}$, and the utilization rate of existing capital $\{u_t\}$ to maximizes its lifetime utility subject to a standard budget constraint:

$$\max \quad E\sum_{t=0}^{\infty}\beta^t\zeta_t\frac{(C_t - bC_{t-1} - \psi h_t^{\theta}S_t)^{1-\sigma}}{1-\sigma} \quad \text{subject to}$$

$$S_t = (C_t - bC_{t-1})^{\gamma}S_{t-1}^{1-\gamma}$$
$$C_t + A_tI_t + G_t = \frac{W_t}{\mu_t}h_t + r_tu_tK_t + P_t$$
$$K_{t+1} = (1-\delta(u_t))K_t + z_t^I I_t\left[1 - \Phi\left(\frac{I_t}{I_{t-1}}\right)\right]$$

Relative to the standard real business cycle model, this model features investment adjustment costs $\Phi(I_t/I_{t-1})$; variable capacity utilization, which increases the return on capital r_tu_t at the cost of increasing its rate of depreciation through $\delta(u_t)$; one period internal habit formation in consumption, controlled by $0 < b < 1$; a potentially low wealth effect on labor supply, when $0 < \gamma < 1$ approaches its lower limit; and monopolistic labor unions, which effectively reduce the wage rate by an amount μ_t each period but rebate profits lump sum to the household through P_t.

Output is produced by a representative firm, which combines capital K_t, labor h_t, and a fixed factor of production L using a (potentially) decreasing returns to scale production function:

$$Y_t = z_t(u_tK_t)^{\alpha_k}(X_th_t)^{\alpha_h}(X_tL)^{1-\alpha_k-\alpha_h}.$$

Market clearing requires that the goods and labor markets clear so that the aggregate resource constraint is satisfied: $C_t + A_tI_t + G_t = Y_t$. The seven fundamental processes capture exogenous variation in permanent and transitory neutral productivity $\{X_t, z_t\}$, permanent and transitory investment-specific productivity $\{A_t, z_t^I\}$, government spending $\{G_t\}$, wage markups $\{\mu_t\}$, and preferences $\{\zeta_t\}$.

B.2 Model of Barsky and Sims (2012)

A representative household chooses consumption $\{C_t\}$, labor supply $\{N_t\}$, and real holdings of riskless one-period bonds $\{B_t\}$ to maximize its lifetime utility subject to a standard budget constraint:

$$\max \quad E\sum_{t=0}^{\infty}\beta^t\left[\ln(C_t - \kappa C_{t-1}) - \frac{N_t^{1+1/\eta}}{1+1/\eta}\right] \quad \text{subject to}$$

$$C_t + B_t = w_t N_t - T_t + (1 + r_{t-1})B_{t-1} + \Pi_t,$$

where r_t is the net nominally risk-free interest rate, w_t is the wage, T_t denotes lump-sum taxes, and Π_t is aggregate profits.

Final goods producers are competitive and take the price of intermediate goods, $P_t(j)$, as given and each have a production function of the form:

$$Y_t = \left[\int_0^1 Y_t(j)^{\frac{\xi-1}{\xi}} \right]^{\frac{\xi}{\xi-1}}$$

Intermediate goods firms are monopolistically competitive and take the demands of final goods firms as given. They each have a production function of the form $Y_t(j) = A_t K_t(j)^{\alpha} N_t(j)^{1-\alpha}$. Each intermediate firm chooses a price for its own good, subject to the constraint that it will only be able to re-optimize its price each period with constant probability $1 - \theta$.

A continuum of capital producers produce new capital (to sell to intermediate firms) according to the production function

$$Y_t^k(\nu) = \phi\left(\frac{I_t(\nu)}{K_t(\nu)}\right) K_t(\nu),$$

where ϕ is an increasing and concave function. The aggregate capital stock evolves according to $K_t = \phi(I_t/K_t)K_{t-1} + (1-\delta)K_{t-1}$, where $0 < \delta < 1$ is the depreciation rate. The aggregate resource constraint is $Y_t = C_t + I_t + G_t$ (ignoring resources lost due to inefficient price dispersion). The monetary authority sets the one-period nominally risk-free rate of return according to a feedback rule of the (log-linear approximate) form:

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)\phi_\pi(\pi_t - \pi^*) + (1 - \rho_i)\phi_y(\Delta Y_t - \Delta Y^*) + \varepsilon_{i,t}.$$

The three fundamental processes capture exogenous variation in permanent neutral productivity $\{A_t\}$, government spending $\{G_t\}$, and monetary policy $\{\varepsilon_{i,t}\}$

B.3 Model of Blanchard, L'Huillier, and Lorenzoni (2013)

A representative household chooses consumption $\{C_t\}$, investment $\{I_t\}$, nominally risk-free bond holdings $\{B_t\}$, and the rate of capital utilization $\{U_t\}$ to maximize its lifetime utility subject to a standard budget constraint:

$$\max \quad E\sum_{t=0}^{\infty}\beta^t\left[\ln(C_t - hC_{t-1}) - \frac{1}{1+\zeta}\int_0^t N_{j,t}^{1+\zeta}dj\right] \quad \text{subject to}$$

$$P_tC_t + P_tI_t + T_t + P_t\mathcal{C}(U_t)\bar{K}_{t-1} = R_{t-1}B_{t-1} + \Upsilon_t + \int_0^1 W_{jt}N_{jt}dj + R_t^kU_t\bar{K}_{t-1},$$
$$\bar{K}_t = (1-\delta)\bar{K}_{t-1} + D_t[1-\mathcal{G}(I_t/I_{t-1})]I_t$$

where P_t is the price level, T_t is a lump sum tax, R_t is the gross nominally risk-free rate, Υ_t is aggregate profits, W_{jt} is the wage of labor type j, R_t^k is the capital rental rate, $0 < \delta < 1$ is the rate of depreciation, $\mathcal{G}(I_t/I_{t-1})$ represents investment adjustment costs, $\mathcal{C}(U_t)$ represents the marginal cost of increasing capacity utilization. It also chooses the wage $\{W_{jt}\}$ for each type of labor subject to the constraint that it will only be able to re-optimize its wage each period with constant probability $1-\theta_w$.

Final goods producers are competitive and take the price of intermediate goods as given, P_{jt}, and each have a production function of the form

$$Y_t = \left[\int_0^1 Y_{jt}^{\frac{1}{1+\mu_{pt}}}dj\right]^{1+\mu_{pt}}.$$

Intermediate goods firms are monopolistically competitive, each with a production function of the form $Y_{jt} = (K_{jt})^{\alpha}(A_tL_{jt})^{1-\alpha}$. Each intermediate firm chooses a price for its own good, subject to a $1-\theta_p$ probability of re-optimization each period.

Labor services are supplied to intermediate goods producers by competitive labor agencies that take wages as given, W_{jt}, and have a production function of the form

$$N_t = \left[\int_0^1 N_{jt}^{\frac{1}{1+\mu_{wt}}}dj\right]^{1+\mu_{wt}}.$$

Market clearing in the final goods market requires that $C_t + I_t + \mathcal{C}(U_t)\bar{K}_{t-1} + G_t = Y_t$, and in the labor market that $\int_0^1 L_{jt}dj = N_t$. Monetary policy follows the rule:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\gamma_\pi\pi_t + \gamma_y\hat{y}_t) + q_t.$$

The six fundamental processes capture exogenous variation in permanent neutral productivity $\{A_t\}$, transitory investment-specific productivity $\{D_t\}$, price markups $\{\mu_{pt}\}$, wage markups $\{\mu_{wt}\}$, government spending $\{G_t\}$, and monetary policy $\{q_t\}$.