On the sources of information about latent variables in DSGE models

Nikolay Iskrev

Abstract

Latent variables are ubiquitous in macroeconomics and structural models are often used to characterize and estimate them from empirical data. This paper addresses the problem of identifying the key sources of information with respect to latent variables in dynamic stochastic general equilibrium models. To that end, I show how to evaluate the information content of a set of observable variables with respect to a given latent variable. The methodology enables researchers to measure and compare the informational contribution of different observables and identify the most informative ones. Thus, it provides a framework for a rigorous treatment of such issues, which are often discussed in an informal manner in the literature. The methodology is illustrated with an assessment of the informational importance of asset prices with respect to news shocks in the business cycle model developed by Schmitt-Grohé and Uribe (2012)

Keywords: DSGE models, Information content, Identification, News Shocks, Asset prices

JEL classification: C32, C51, C52, E32

Bank of Portugal, Research in Economics and Mathematics (REM), and Research Unit on Complexity and Economics (UECE). UECE is supported by Fundação para a Ciência e a Tecnologia.
Av. Almirante Reis 71, 1150-012, Lisboa, Portugal, +351213130006, nikolay.iskrev@bportugal.pt
Acknowledgments: I am grateful to the editor, Eric Leeper, an associate editor, and two anonymous referees for their for their detailed comments and suggestions that substantially improved the paper. I would also like to thank Stephanie Schmitt-Grohé, Ellen McGrattan, Marek Jarociński, Mark Watson, and participants in the 19th Central Bank Macroeconomic Modelling Workshop hosted by Central Bank of Armenia for useful conversations and comments. The views expressed in this publication are those of its author and do not necessarily reflect the views of Banco de Portugal or the Eurosystem.

1 Introduction

Latent variables are ubiquitous in modern macroeconomics. Examples include endogenously determined quantities, such as potential output, permanent income, and natural interest rate, as well as a wide range of exogenous shocks hypothesized by different economic theories to be the primary sources of aggregate fluctuations. Such variables are treated as unobservable because they represent purely theoretical concepts and nothing we see in the real world directly matches them. Models based on economic theory, which characterize relationships between unobserved variables and observable quantities, are required to measure the former using empirical data. Fully articulated dynamic stochastic general equilibrium (DSGE) models are particularly suitable for that purpose as they provide a framework for modeling the structure linking multiple potentially observable variables on one hand, and latent ones, on the other, in an internally consistent manner.

Estimated latent variables play a crucial role when models are used to give a structural interpretation of the observed movements in key macroeconomic aggregates. The potential of DSGE models to provide a plausible account of economic fluctuations is often invoked as a major advantage over using reduced-form models, which often yield a superior fit to the data. In practice, this is achieved by conducting historical decompositions of both observed and unobserved endogenous variables in terms of the underlying exogenous forces driving them. Such decompositions are only credible if the latent variables of interest are accurately estimated, i.e. if the observed data is sufficiently informative about them.

The purpose of this paper is to show how to assess DSGE models' implications for the contribution of information of observable variables with respect to latent ones. Such issues are frequently discussed in a heuristic fashion in the literature. My aim here is to describe how to approach them in a formal manner. Intuitively, information can be interpreted as the reduction of uncertainty about an unknown quantity. This is made precise by using well-established measures of uncertainty, mutual information and information gains developed in information theory. The transfer of information between variables is quantified by comparing different probability distributions, e.g. distributions of a given shock conditional on nested sets of observables. The required distributions are completely characterized by the underlying structural model, and in the class of linearized Gaussian DSGE models those quantities are available in closed form.

One practical application of the methodology is to reveal what type of data are

required to best identify a given structural shock. As noted earlier, this is often a key requirement for estimated models to provide credible interpretations of past and present economic developments. Understanding where information about different shocks comes from is not always a simple matter, however, especially in cases where new and less well understood types of shocks are incorporated into otherwise standard macroeconomic models. The estimation of such models typically involves, in addition to standard macroeconomic series, other variables, which the authors believe are informative about the shocks they introduce. For instance, data on corporate cash flows, loans, equity, and bond yields have been used to estimate models with financial shocks (see e.g. Gilchrist et al. (2009a), Kaihatsu and Kurozumi (2014) and Ajello (2016)). Financial variables are also used in Christiano et al. (2014) to help identify risk shocks in their model. Other recent examples in the same vein are Ilut and Schneider (2014), who use data on survey forecasts dispersion to identify confidence shocks, and Liu et al. (2013) who use land price data in the estimation of a model featuring housing demand shocks.

To illustrate the proposed methodology, I examine the informational relevance of asset prices in a DSGE model with news shocks. The existing literature on news-driven business cycles is divided on the subject of whether such models can be estimated using only standard macro variables, or if other types of data are required to identify news shocks. Among the studies using only standard macro variables are Fujiwara et al. (2011), Milani and Treadwell (2012), and Schmitt-Grohé and Uribe (2012). Asset prices are used to estimate models with news by Davis (2007), Avdjiev (2016), Görtz and Tsoukalas (2017), and Bretscher et al. (2019), while Hirose and Kurozumi (2012), Milani and Rajbhandari (2012), and Miyamoto and Nguyen (2019) use survey data on expectations. The rationale for using asset prices in particular is based on the intuition that, due to the existence of various real and nominal rigidities in the economy, macro variables are far less sensitive to news than asset prices. Consequently, asset price variables are perceived to be more informative about news shocks than the variables commonly used to estimate models without news. Some empirical evidence to support this view can be found in the structural VAR literature, where variables such as stock prices often play a central role for the identification of news shocks. For instance, Görtz and Tsoukalas (2017) invoke the findings of Gilchrist et al. (2009b) and Gilchrist and Zakrajšek (2012) to motivate the use of corporate bond spreads in the estimation of their model. Similarly, Bretscher et al. (2019) cite the work of Beaudry and Portier (2006) and Kurmann and Otrok (2013), who found stock prices and the slope of the term structure of interest rates to be informative

about news shocks, to motivate the inclusion of these variables in their estimation.

The methodology presented here can be used to determine if the intuition that asset prices are more informative about news shocks than macro variables is correct in the context of a given model. I consider the model estimated in Schmitt-Grohé and Uribe (2012) and study the information contributed by two asset prices – the value of the representative firm and the risk-free real interest rate. The results suggest relatively small informational value of observing either variable. While including asset prices as observables increases the amount of information about some news shocks, their marginal contributions are comparable to the contributions of non-asset price variables, such as hours worked, total factor productivity, or the relative price of investment. The only news shocks with respect to which asset prices, specifically the risk-free interest rate, are found to be more informative than any other variable are news about the stationary neutral productivity shock.

While the primary focus of this article is on identifying the key sources of information with respect to latent variables, in some cases it may also be useful to know where information about particular model parameters, related to those variables, come from. For instance, in the context of news-driven models, the claim that asset prices are especially useful for identifying news shocks may refer to information these variables contribute with respect to parameters characterizing the marginal distributions of news shocks. To evaluate the information content of observables with respect to parameters, I compute measures of efficiency gains that compare the values of the Cramér-Rao lower bounds conditional on different sets of observed variables. Again, the results show that while including asset prices as observables would lead to significant efficiency gains with respect to the news shock parameters, the gains are similar in size to those due to standard macroeconomic variables, such as hours worked and total factor productivity.

The remainder of the article is organized as follows. Section 2 discusses the relationship between this paper and the existing literature. Section 3 gives an overview of the relevant information-theoretic concepts, and defines measures of information gains with respect to latent variables, and efficiency gains with respect to parameters. The proposed measures are then applied, in Section 4, to evaluate the information content of asset prices in a DSGE model with news shocks. Section 5 concludes.

2 Relationship to prior literature

In terms of methodology, this paper is related to a large literature on measuring the relative importance of variables in scientific models. A common application of this type of analysis is to determine the relative importance of individual regressors in explaining the behavior of response variables (Kruskal (1987)). The use of information-theoretic measures in that context dates back at least to Theil (1987), who uses a decomposition of the Gaussian mutual information to quantify the contribution of independent explanatory variables in multivariate regressions.[1] A comprehensive treatment of the subject from an information theory perspective is given in Retzer et al. (2009), who characterize the importance of variables by the extent to which their use reduces uncertainty about predicting the response variable. Another important area of application is the study of causal relationships in the analysis of time series. Following the seminal work of Granger (1969), the notion of causality has been associated with the question of whether knowledge of past values of one time series helps improve the prediction of another.

While most of the early work on this topic focused on how to test for the existence and direction of causality, Geweke (1982, 1984) show how to quantify the strength of causal influence. Geweke's measures are based on the magnitude of the reduction of forecast uncertainty, measured by the mean square forecast errors of the predicted variable, due to using past values of the causal variable. In that sense, measuring Granger causality can be interpreted as quantifying the contribution of information by observed variables – past observations of the cause variable, with respect to unobserved ones – the future values of the predicted variable, conditional on a set of other observed variables – the past values of the predicted variable.[2] This is precisely the meaning of conditional mutual information, and, as Barnett et al. (2009) show, when the joint distribution of the variables is Gaussian, Geweke's measures of strength of causality are equivalent to the "transfer entropy" of Schreiber (2000), which is an information-theoretic measure of the transfer of information between two stochastic processes.[3] Extensions to non-linear and non-Gaussian models involve replacing the forecast error variances with entropic

[1] See also Theil and Chung (1988) where the analysis is extended to systems of simultaneous equations, and Soofi (1992), who applies the same ideas to determine the relative importance of predictors in logit models.

[2] In his Nobel prize acceptance lecture Granger defined causality as follows: (1) The cause occurs before the effect; (2) The cause contains information about the effect that is unique, and is in no other variable.

[3] See also Pourahmadi and Soofi (2000) who use conditional mutual information to quantify the information worth of past observations for predicting future values of univariate time series.

measures of uncertainty (see Amblard and Michel (2011)).[4]

Instead of strength of causality, the purpose of the measures presented in this paper is to quantify the amount of information that realizations of observed variables contribute with respect to contemporaneous but unobserved realizations of some endogenous model variables or exogenous shocks. Since mathematically there is no difference between unobserved future realizations of observed variables and unobserved contemporaneous realizations of latent variables, the proposed measures of information gains are analogous, with minor modifications, to the Granger causality strength measures of Geweke (1982, 1984) in the case of linearized Gaussian DSGE models, and to non-linear Granger causality measures in the general case.

The paper is also related to a growing literature on the feasibility of recovering structural shocks using reduced form models. Building upon the work of Hansen and Sargent (1980, 1991) and Lippi and Reichlin (1993, 1994), most of the research on this topic has focused on the issue of invertibility (or fundamentalness) in structural vector autoregressions, i.e. whether shocks from general equilibrium models can be recovered from the residuals of VARs (see Alessi et al. (2011) and Giacomini (2013) for useful overviews of this literature). Conditions for invertibility are discussed in Fernandez-Villaverde et al. (2007), Ravenna (2007), Franchi and Vidotto (2013), Franchi and Paruolo (2015)), while Giannone and Reichlin (2006) and Forni and Gambetti (2014) discuss how to test for lack of invertibility of structural VARs. Invertibility issues that are specific to DSGE models with news shocks are discussed in Leeper et al. (2013) and Blanchard et al. (2013). More recently, Soccorsi (2016) and Forni et al. (2016) proposed measures of the degree of non-invertibility, which quantify the discrepancies between true shocks and shocks obtained using non-fundamental VARs.[5] In another recent paper Chahrour and Jurado (2017) draw a distinction between invertibility on one hand, and what they call "recoverability" on the other, defining the latter as the feasibility of recovering structural shocks from all available observations, not only past and present ones. They argue that recoverability is often what matters in applied research, and present a necessary and sufficient condition one can use to check if shocks in linear models are recoverable.

[4]A concept related to Granger causality is that of Granger causal priority, which Jarociński and Maćkowiak (2017) utilize in a recent article to show how to determine the relevant variables to be included in Bayesian vector autoregressions.

[5]Simulation evidence that non-invertible VARs may in some cases produce good approximations of the true structural shocks are provided in Sims (2012) and Beaudry et al. (2015).

Similar to that literature, the analysis in the present paper may be used to determine whether the shocks in a DSGE model can be recovered from a set of observed variables. And, like in Chahrour and Jurado (2017), the analysis is based on information contained in all available observations. Furthermore, similar to Soccorsi (2016) and particularly Forni et al. (2016), a measure is provided of the degree to which any individual shock, or an endogenous latent variable, can be recovered. In particular, the proposed measures of information gains are defined with respect to a particular unobserved variable and show how much of the prior uncertainty about it is removed due to observing a given set of model variables. An important difference with the invertibility literature is that the analysis here is based on the true data generating process characterized by a structural model, and not on approximations of it, such as a VAR. The proposed information gain measures are, in their general form, meaningful and useful when applied to non-linear DSGE models, while the invertibility conditions and measures in the existing literature are specific to linearized models. In the context of linearized DSGE models, the information gain measures could be interpreted as upper bounds on the amount of information about a shock (or the degree of information sufficiency in the terminology of Forni et al. (2016)) available in a VAR.

More importantly, while the existing research on invertibility is concerned with the usefulness of VAR–based tools for empirical validation of structural models, the purpose of the analysis presented here is to understand the properties of DSGE models in terms of how information transfers between observed and unobserved model variables. Therefore, identifying the main sources of information is of greater interest than what the total amount of information about a given shock is. To that end, I define and apply measures of conditional information gains that quantify the amount of additional information contributed by a variable or several variables, given the information contained in another set of observed variables. As the analysis of the model considered in the application section shows, the conclusions one draws may be very different depending on what the conditional variables are. For instance, asset prices are found to be unconditionally very informative with respect to wage markup news shocks in the model of Schmitt-Grohé and Uribe (2012), but conditional on observing other macro variables, the information gains are small. At the same time, asset prices may be conditionally quite informative about certain productivity news shocks even though the unconditional information gains are close to zero. These findings are a reflection of the fact that information contained in different variables is not necessarily independent and could be overlapping

in some cases while complementary in others. A somewhat extreme example of this phenomenon is the finding that output growth is informationally completely redundant in the model of Schmitt-Grohé and Uribe (2012). In general, information contained in different variables tends to be partially redundant with respect to some latent variables and complementary with respect to others. A novel measure of pairwise information complementarity is introduced in the paper and used to determine the sign and assess the degree of complementarity among observed variables with respect to unobserved ones. In the application section, the measure is used to clarify the nature of the interactions between asset prices and other macroeconomic variables in terms of information they convey with respect to news shocks in the Schmitt-Grohé and Uribe (2012) model.

3 Measures of information and information gains

A DSGE model completely characterizes the joint probability distribution of a $n_{\boldsymbol{y}}$ vector of *observed* endogenous variables $\boldsymbol{y}$, and a $n_{\boldsymbol{z}}$ vector of *unobserved* endogenous variables and exogenous shocks $\boldsymbol{z}$. Note that in practice the dimension of each of these vectors is a function of a sample size T. For notational simplicity I suppress the dependence on T throughout this section unless it is necessary to make it explicit. The joint distribution function of $\boldsymbol{y}$ and $\boldsymbol{z}$ is parameterized in terms of a $n_{\boldsymbol{\theta}}$ vector of structural parameters $\boldsymbol{\theta}$, characterizing technology, preferences, and the properties of the exogenous variables. Both $\boldsymbol{\theta}$ and $\boldsymbol{z}$ are typically unknown and unobserved, and the only source of empirical information about them are the measurements of $\boldsymbol{y}$. The purpose of this section is to show how to quantify the amount of information contained in a sample of data, and how to evaluate the contributions of individual observed variables.

One way to approach these questions would be to adopt a Bayesian perspective and treat $\boldsymbol{z}$ as part of the parameter vector to be estimated. Then, the amount of information provided by a sample would be with respect to $(\boldsymbol{\theta}, \boldsymbol{z})$ jointly. While conceptually feasible, this approach would be very challenging in practice in the present context, given the large dimension of $\boldsymbol{z}$ and the complicated form of the conditional distribution of $(\boldsymbol{\theta}, \boldsymbol{z})$ given $\boldsymbol{y}$. Therefore, in most of this section I treat $\boldsymbol{\theta}$ as known and measure sample information and information gains about $\boldsymbol{z}$ conditional on $\boldsymbol{\theta}$. At the end of the section I discuss the issue of measuring sample information about $\boldsymbol{\theta}$.

3.1 Information about latent variables

A well-established measure of information about random variables is the information-theoretic entropy introduced by Shannon (1948). Entropy is a measure of the uncertainty associated with a random variable, and the amount of information about that variable is measured as the reduction in uncertainty, i.e. entropy, relative to some base distribution. Specifically, let $f(\boldsymbol{z})$ be the probability density function of $\boldsymbol{z}$. For notational simplicity throughout this subsection I suppress the dependence on $\boldsymbol{\theta}$. The entropy $\mathrm{H}(\boldsymbol{z})$ of $f(\boldsymbol{z})$ is defined as

$$\mathrm{H}(\boldsymbol{z}) = -\int f(\boldsymbol{z}) \ln\left(f(\boldsymbol{z})\right) d\boldsymbol{z} = -\,\mathrm{E} \ln f(\boldsymbol{z}). \tag{3.1}$$

Similarly, if $f(\boldsymbol{y}, \boldsymbol{z})$ is the joint probability density function of $\boldsymbol{y}$ and $\boldsymbol{z}$, the joint entropy $\mathrm{H}(\boldsymbol{y}, \boldsymbol{z})$ of $f(\boldsymbol{y}, \boldsymbol{z})$ is defined as

$$\mathrm{H}(\boldsymbol{y}, \boldsymbol{z}) = -\int f(\boldsymbol{y}, \boldsymbol{z}) \ln\left(f(\boldsymbol{y}, \boldsymbol{z})\right) d\boldsymbol{y} d\boldsymbol{z} = -\,\mathrm{E} \ln f(\boldsymbol{y}, \boldsymbol{z}) \tag{3.2}$$

The difference between joint and marginal entropies

$$\mathrm{H}(\boldsymbol{z}|\boldsymbol{y}) = \mathrm{H}(\boldsymbol{y}, \boldsymbol{z}) - \mathrm{H}(\boldsymbol{y}) \tag{3.3}$$

defines the conditional entropy of $\boldsymbol{z}$ given $\boldsymbol{y}$. It measures the amount of uncertainty about $\boldsymbol{z}$ that remains once $\boldsymbol{y}$ is observed. Note that $\mathrm{H}(\boldsymbol{z}|\boldsymbol{y})$ can be computed as in (3.1) using the conditional density $f(\boldsymbol{z}|\boldsymbol{y})$ of $\boldsymbol{z}$ given $\boldsymbol{y}$. It can be shown (see for instance Granger and Lin (1994)) that $\mathrm{H}(\boldsymbol{z}) \geq \mathrm{H}(\boldsymbol{z}|\boldsymbol{y})$ with equality if and only if $f(\boldsymbol{y}, \boldsymbol{z}) = f(\boldsymbol{y}) f(\boldsymbol{z})$. Hence, unless $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent, observing $\boldsymbol{y}$ provides information about $\boldsymbol{z}$. The amount of uncertainty about $\boldsymbol{z}$ that is removed by observing $\boldsymbol{y}$ is known as the mutual information of $\boldsymbol{y}$ and $\boldsymbol{z}$, i.e.[6]

$$\mathrm{I}(\boldsymbol{y}, \boldsymbol{z}) = \mathrm{H}(\boldsymbol{z}) - \mathrm{H}(\boldsymbol{z}|\boldsymbol{y}). \tag{3.4}$$

$\mathrm{I}(\boldsymbol{y}, \boldsymbol{z})$ is a measure of information in the sense that it quantifies the expected reduction in uncertainty about one of the variables due to observing the other one. From $\mathrm{H}(\boldsymbol{z}) \geq$

[6] Mutual information is defined as $\mathrm{I}(\boldsymbol{y}, \boldsymbol{z}) = \int f(\boldsymbol{y}, \boldsymbol{z}) \ln \dfrac{f(\boldsymbol{y}, \boldsymbol{z})}{f(\boldsymbol{y}) f(\boldsymbol{z})} d\boldsymbol{y} d\boldsymbol{z}$ and measures the distance between the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{z}$ and the distribution when the variables are independent. See Cover and Thomas (2006) for more details on the properties of entropy and mutual information.

$\mathrm{H}(\boldsymbol{z}|\boldsymbol{y})$ it follows that mutual information is positive unless $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent in which case it is zero. On the other hand, if the variables are perfectly dependent i.e. there exists a one-to-one function g such that $\boldsymbol{z} = g(\boldsymbol{y})$, observing $\boldsymbol{y}$ is equivalent to observing $\boldsymbol{z}$. In that case $\mathrm{I}(\boldsymbol{y}, \boldsymbol{z}) = \infty$ (see Granger and Lin (1994, Theorem 2)). It is common in practice to normalize the measure to be in the interval $[0,\ 1]$. For instance, Joe (1989) proposed the following transformation

$$\mathrm{I}^*(\boldsymbol{y}, \boldsymbol{z}) = 1 - \exp\left(-2\mathrm{I}(\boldsymbol{y}, \boldsymbol{z})\right) \tag{3.5}$$

as a generalized measure of dependence between two or more random variables. The same transformation is used in Granger and Lin (1994) as a criterion for determining the number of significant lags in nonlinear time series models. The reason why the particular form in (3.5) is chosen is that, for a bivariate Gaussian distribution, $\mathrm{I}^*(\boldsymbol{y}, \boldsymbol{z}) = \rho^2$, where ρ is the linear correlation coefficient between $\boldsymbol{y}$ and $\boldsymbol{z}$. Furthermore, when $\boldsymbol{y}$ and $\boldsymbol{z}$ are jointly Gaussian, the transformation in (3.5) results in the following expression[7]

$$\mathrm{I}^*(\boldsymbol{y}, \boldsymbol{z}) = \frac{|\boldsymbol{\Sigma}_{\boldsymbol{z}}| - |\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{y}}|}{|\boldsymbol{\Sigma}_{\boldsymbol{z}}|}, \tag{3.6}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{z}}$ is the covariance matrix of the marginal probability density of $\boldsymbol{z}$, and $\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{y}}$ is the covariance matrix of the conditional probability density of $\boldsymbol{z}$ given $\boldsymbol{y}$. Hence, for Gaussian distributions, $\mathrm{I}^*(\boldsymbol{y}, \boldsymbol{z})$ measures the reduction in the generalized variance (Wilks (1932)) of vector $\boldsymbol{z}$ due to observing vector $\boldsymbol{y}$, as a fraction of the unconditional generalized variance of $\boldsymbol{z}$. However, as Peña and Rodríguez (2003) and others have noted, the generalized variance is not a dimensionless measure of the uncertainty of a (Gaussian) random vector. For instance, if $\boldsymbol{\Sigma}_{\boldsymbol{z}}$ is a $n_{\boldsymbol{z}} \times n_{\boldsymbol{z}}$ diagonal matrix with $\sigma^2 < 1$ on the diagonal, $|\boldsymbol{\Sigma}_{\boldsymbol{z}}| = \sigma^{2n_{\boldsymbol{z}}}$, implying exponential decline of uncertainty as the dimension of $\boldsymbol{z}$ grows. A dimensionless measure of variability proposed in Peña and Linde (2007) is the effective variance $\mathrm{V}_e(\boldsymbol{z})$, defined as

$$\mathrm{V}_e(\boldsymbol{z}) = |\boldsymbol{\Sigma}_{\boldsymbol{z}}|^{1/n_{\boldsymbol{z}}}. \tag{3.7}$$

[7] This follows from the result that the entropy of a $n_{\boldsymbol{v}}$–dimensional Gaussian variable $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{v}})$ is $\mathrm{H}(\boldsymbol{v}) = 0.5\left(\ln(2\pi e)^{n_v} + \ln|\boldsymbol{\Sigma}_{\boldsymbol{v}}|\right)$. Therefore, the mutual information of $\boldsymbol{y}$ and $\boldsymbol{z}$ is $\mathrm{I}(\boldsymbol{y}, \boldsymbol{z}) = \mathrm{H}(\boldsymbol{z}) - \mathrm{H}(\boldsymbol{z}|\boldsymbol{y}) = .5 \ln\left(\frac{|\boldsymbol{\Sigma}_{\boldsymbol{z}}|}{|\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{y}}|}\right)$.

When the elements of $\boldsymbol{z}$ are independent, i.e. $\boldsymbol{\Sigma}_{\boldsymbol{z}}$ is diagonal, the effective variance is equal to the geometric average of the variances of the elements of $\boldsymbol{z}$. If, in addition, the variances are identical, the effective variance is equal to that common variance, and is therefore independent of T. In the general case, $\mathrm{V}_e(\boldsymbol{x})$ is equal to the geometric average of the eigenvalues of $\boldsymbol{\Sigma}$. Adopting the effective variance as a scalar measure of the uncertainty associated with Gaussian distributions yields the following measure of the information gained about $\boldsymbol{z}$ from observing $\boldsymbol{y}$:

$$\mathrm{IG}_{\boldsymbol{z}}(\boldsymbol{y}) = \left(\frac{|\boldsymbol{\Sigma}_{\boldsymbol{z}}|^{1/n_{\boldsymbol{z}}} - |\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{y}}|^{1/n_{\boldsymbol{z}}}}{|\boldsymbol{\Sigma}_{\boldsymbol{z}}|^{1/n_{\boldsymbol{z}}}} \right) \times 100. \tag{3.8}$$

The interpretation of $\mathrm{IG}_{\boldsymbol{z}}(\boldsymbol{y})$ is the following: it measures the reduction in uncertainty about vector $\boldsymbol{z}$ due to observing vector $\boldsymbol{y}$, as a percent of the unconditional (prior) uncertainty about $\boldsymbol{z}$.

The measure in (3.8) can be generalized for non-Gaussian distribution by noting that $\mathrm{V}_e(\boldsymbol{x})$ is equal to a particular transformation of the entropy $\mathrm{H}(\boldsymbol{z})$ when $\boldsymbol{z}$ is Gaussian. Specifically, Shannon (1948) defined the entropy power $\mathrm{N}(\boldsymbol{z})$ of a vector $\boldsymbol{z}$ with entropy $\mathrm{H}(\boldsymbol{z})$ to be

$$\mathrm{N}(\boldsymbol{z}) = \frac{1}{2\pi e} \exp\left(\frac{2}{n_{\boldsymbol{z}}} \mathrm{H}(\boldsymbol{z}) \right), \tag{3.9}$$

which for $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{z}}, \boldsymbol{\Sigma}_{\boldsymbol{z}})$ implies $\mathrm{N}(\boldsymbol{z}) = |\boldsymbol{\Sigma}_{\boldsymbol{z}}|^{1/n_{\boldsymbol{z}}}$. Similarly, the conditional entropy power of $\boldsymbol{z}$ given $\boldsymbol{y}$ is $\mathrm{N}(\boldsymbol{z}|\boldsymbol{y}) = |\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{y}}|^{1/n_{\boldsymbol{z}}}$. Note that, unlike entropy which can be negative (for continuous variables), entropy power is always non-negative, and is therefore a more appealing measure of uncertainty. Thus, $\mathrm{IG}_{\boldsymbol{z}}(\boldsymbol{y})$ can be defined for non-Gaussian distribution as in (3.8), replacing the effective variance with entropy power. It can also be expressed in terms of mutual information using the following transformation:

$$\mathrm{IG}_{\boldsymbol{z}}(\boldsymbol{y}) = \left(1 - \exp\left(-\frac{2}{n_{\boldsymbol{z}}} \mathrm{I}(\boldsymbol{y}, \boldsymbol{z}) \right) \right) \times 100. \tag{3.10}$$

Hence $\mathrm{IG}_{\boldsymbol{z}}(\boldsymbol{y})$ is a simple modification of the transformation in (3.5) that allows information gains to be compared for vectors of different dimensions.

In the context of DSGE models, we are often interested in the information content of one or more observed variables with respect to a particular latent variable. Hence, the relevant information gain measure is of the form $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i)$, where $\boldsymbol{z}^j$ is a $n_{\boldsymbol{z}^j}$ sub-vector

of $\boldsymbol{z}$ containing the realization of the latent variable we are interested in, and $\boldsymbol{y}^i$ is a $n_{\boldsymbol{y}^i}$ sub-vector of $\boldsymbol{y}$ containing the observations of the variable or variables whose information content we want to assess. The required quantities, i.e. entropy (3.1) and mutual information (3.4), are obtained in exactly the same way as before, replacing the joint distributions with their marginal counterparts. Furthermore, now we can distinguish between conditional and unconditional information gains from knowing $\boldsymbol{y}^i$ with respect to $\boldsymbol{z}^j$. The unconditional information gain is given as before by $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i)$ and measures the reduction in uncertainty about $\boldsymbol{z}^j$ due to observing $\boldsymbol{y}^i$ relative to observing no data at all. It is often more interesting to know the marginal contribution of $\boldsymbol{y}^i$, given the information about $\boldsymbol{z}^j$ contained in other observed variables. One way to define a conditional information gain of $\boldsymbol{y}^i$ with respect to $\boldsymbol{z}^j$, given $\overline{\boldsymbol{y}}^i = \boldsymbol{y} \setminus \boldsymbol{y}^i$ is to replace the mutual information $\mathrm{I}(\boldsymbol{y}^i, \boldsymbol{z}^j)$ in (3.10) with the conditional mutual information $\mathrm{I}(\boldsymbol{y}^i, \boldsymbol{z}^j|\overline{\boldsymbol{y}}^i) = \mathrm{H}(\boldsymbol{z}^j|\overline{\boldsymbol{y}}^i) - \mathrm{H}(\boldsymbol{z}^j|\boldsymbol{y})$; this would tell us how much of the uncertainty about $\boldsymbol{z}^j$ that remains after $\overline{\boldsymbol{y}}^i$ is observed is removed by observing also $\boldsymbol{y}^i$.[8] Note, however, that the gains would be relative to the conditional uncertainty about $\boldsymbol{z}^j$ given $\overline{\boldsymbol{y}}^i$. Therefore, that measure is not comparable to $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i)$ in (3.10), which is in terms of percent of the unconditional uncertainty about $\boldsymbol{z}^j$. A conditional measure, comparable to (3.8) in the Gaussian case, can be defined as

$$\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i|\overline{\boldsymbol{y}}^i) = \left(\frac{|\boldsymbol{\Sigma}_{\boldsymbol{z}^j|\overline{\boldsymbol{y}}^i}|^{1/n_{\boldsymbol{z}^j}} - |\boldsymbol{\Sigma}_{\boldsymbol{z}^j|\boldsymbol{y}}|^{1/n_{\boldsymbol{z}^j}}}{|\boldsymbol{\Sigma}_{\boldsymbol{z}^j}|^{1/n_{\boldsymbol{z}^j}}} \right) \times 100. \tag{3.11}$$

The interpretation of $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i|\overline{\boldsymbol{y}}^i)$ is the following: it shows the amount of uncertainty about $\boldsymbol{z}^j$ left after observing $\overline{\boldsymbol{y}}^i$ that is removed by observing also $\boldsymbol{y}^i$, as a percent of the unconditional uncertainty about $\boldsymbol{z}^j$. As before, for non-Gaussian distributions the (conditional) effective variances are replaced with the respective (conditional) entropy powers.

Note that, while the information gain measures in (3.8) and (3.11) are defined with respect to the T-dimensional vector of all realizations of a given latent variable, we can similarly compute information gains with respect to smaller subsets of realizations. For instance, we can define information gains with respect to an individual realization of the latent variable at a given point in time. In that case we use the marginal conditional distributions of that realization for different conditioning variables. If the distribution is Gaussian, the conditional variances that appear in the information gain measures may be

[8] The notation $A = B \setminus C$ is used to define A as the subset of B that excludes the set C.

obtained by using the Kalman smoother. Note, however, that the Kalman smoother only provides the diagonal elements of the conditional covariance matrices in (3.8) and (3.11). In general, these matrices have non-zero off-diagonal entries since the elements of $\boldsymbol{z}^j$ may be conditionally dependent even if the latent variable itself is an i.i.d. process. Therefore, the standard Kalman smoother cannot be used to evaluate $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i)$ or $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}^i|\bar{\boldsymbol{y}}^i)$ when $\boldsymbol{z}^j$ contains two or more realizations of a latent variable. Instead, for linearized Gaussian models, like the one analyzed in the next section, I use the fact that $\boldsymbol{y}$ can be expressed as an affine function of $\boldsymbol{z}$, which implies that the conditional distribution of any subset of $\boldsymbol{z}$ given $\boldsymbol{y}$ (or any subset of $\boldsymbol{y}$) is also Gaussian. Consequently, the required covariances matrices are simple to obtain using the familiar formulas for moments of conditional Gaussian distributions. For more details, see Section A.1 of the Online Appendix.[9]

The use of the information gain measures presented above can be summarized as follows: the unconditional measure (3.8) tells us how informative a set of observed variables is as a whole with respect to a given unobserved endogenous variable or exogenous shock. If $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}) \approx 0$ the information about $\boldsymbol{z}^j$ after observing $\boldsymbol{y}$ is nearly the same as prior to observing any data. For instance, saying that standard macroeconomic variables are not very informative about news shocks can be expressed as the unconditional information gains of such variables (as a set) with respect to news shocks being close to zero. On the other hand, if $\mathrm{IG}_{\boldsymbol{z}^j}(\boldsymbol{y}) = 100$, observing $\boldsymbol{y}$ is sufficient to completely recover the realizations of the variable represented by $\boldsymbol{z}^j$. The conditional information gain measure (3.11) can be used to determine how much of the overall information content of $\boldsymbol{y}$ is contributed by each individual variable or subsets of variables. Therefore, the claim that asset prices contribute a lot of additional information about news shocks can be verified by showing that the information gains of asset prices with respect to news shocks conditional on standard macroeconomic variables are large. In general, by comparing conditional information gains, one could rank observed variables in terms of their relative informativeness with respect to each latent variable.

As mentioned in the Introduction, in some cases questions about the informational importance of observed variables may refer to information about unknown parameters related to a given latent variable, instead of information about the realizations of

[9] An example of a situation where the uncertainty for more than two but less than T realizations is of interest is having a latent variable whose realizations can be recovered fully after some point in time $T_1 < T$. This could be due to uncertainty about the initial state that propagates through the first few periods. In that case, the conditional variance of the first T_1 realizations is positive while that of the last $T - T_1$ is zero.

that variable. Informative variables in that sense are those which, if observed, would substantially reduce the estimation uncertainty of the parameters in question. In the remainder of this section I discuss how to assess the relative importance of the observed variables with respect to model parameters.

3.2 Information about parameters

The purpose of this section is to show how to evaluate the amount of information observed variables contribute with respect to the vector $\boldsymbol{\theta}$ as a whole, as well as individual parameters. I approach this as a missing data problem (see e.g. Dempster et al. (1977) and Palm and Nijman (1984)), and compare the expected information content of complete and incomplete samples. In the present context, having a complete sample means observing an $n_{\boldsymbol{y}_T}$ vector $\boldsymbol{y}_T$, while having an incomplete sample means observing $\overline{\boldsymbol{y}}_T^i$, that is, all variables except the one indexed by i. Intuitively, the distribution of the incomplete sample is less informative than the distribution of the complete sample in the sense that the uncertainty about $\boldsymbol{\theta}$ is reduced to a lesser extent as a consequence of observing $\overline{\boldsymbol{y}}_T^i$ compared to observing $\boldsymbol{y}_T$ (see Rao (2002, p.331)).[10] A standard measure of the expected amount of information contained in a distribution is the Fisher information matrix (hereafter denoted by FIM). Asymptotically, i.e. as T tends to infinity, the inverse of the FIM is equal to the covariance matrix of the distribution of the ML estimator of $\boldsymbol{\theta}$. Hence, the expected loss of information can be measured by comparing the asymptotic variances of MLE using complete and incomplete samples. Furthermore, by the Cramér-Rao theorem the inverse of FIM is a lower bound on the covariance matrix of any unbiased estimator of $\boldsymbol{\theta}$. Thus, the loss of information can also be assessed as a function of the sample size, by measuring the differences between Cramér-Rao lower bounds (hereafter denoted by CRLB) with complete and incomplete samples.

For consistency with the previous section, I evaluate the contributions of observables in terms of efficiency gains, i.e. the reduction in expected estimation uncertainty about parameters when a variable (or a group of variables) is observed, relative to when it is not observed. Specifically, let $\boldsymbol{\Omega}_{\boldsymbol{\theta}}(\boldsymbol{y})$ and $\boldsymbol{\Omega}_{\boldsymbol{\theta}}(\overline{\boldsymbol{y}}^i)$ be the (asymptotic or finite-sample) CRLBs for $\boldsymbol{\theta}$ associated with the complete and incomplete samples. Note that, for large T, the MLE of $\boldsymbol{\theta}$ is an approximately Gaussian random vector with mean equal to the true value of $\boldsymbol{\theta}$, and covariance matrix given by the CRLB. Hence, $|\boldsymbol{\Omega}_{\boldsymbol{\theta}}|^{1/n_{\boldsymbol{\theta}}}$ can

[10] See Meng and Xie (2014) for an interesting discussion of why this is true for likelihood based estimation approaches, but not in general.

be interpreted as a large-sample approximation of the entropy power of the MLE of $\boldsymbol{\theta}$. Following the discussion in the previous section, I define the efficiency gain with respect to $\boldsymbol{\theta}$ from observing $\boldsymbol{y}^i$ relative to observing $\bar{\boldsymbol{y}}^i$ as follows:

$$\mathrm{EG}_{\boldsymbol{\theta}}(\boldsymbol{y}^i|\bar{\boldsymbol{y}}^i) = \left(\frac{|\boldsymbol{\Omega}_{\boldsymbol{\theta}}(\bar{\boldsymbol{y}}^i)|^{1/n_{\boldsymbol{\theta}}} - |\boldsymbol{\Omega}_{\boldsymbol{\theta}}(\boldsymbol{y})|^{1/n_{\boldsymbol{\theta}}}}{|\boldsymbol{\Omega}_{\boldsymbol{\theta}}(\bar{\boldsymbol{y}}^i)|^{1/n_{\boldsymbol{\theta}}}} \right) \times 100. \tag{3.12}$$

The interpretation is similar to that of the information gains in the previous section. $\mathrm{EG}_{\boldsymbol{\theta}}(\boldsymbol{y}^i|\bar{\boldsymbol{y}}^i)$ shows the increase in (asymptotic) efficiency of the MLE of $\boldsymbol{\theta}$ due to observing $\boldsymbol{y}^i$ as a percent of the estimation efficiency when only $\bar{\boldsymbol{y}}^i$ is observed. I use *efficiency gain* instead of *information gain* to emphasize the fact that, unlike the previous section, the gains here are in terms of the uncertainty associated with the distribution of the estimator of $\boldsymbol{\theta}$, instead of the parameter itself, which is non-random. For details on how to evaluate the FIM for linearized Gaussian models, like the one analyzed in the next section, see Section A.2 of the Online Appendix.

Wei (1978a,b) uses a similar measure to assess the information loss due to aggregation of time series. The only difference is that he does not exponentiate the determinants of the asymptotic covariance matrices of MLE for aggregated and disaggregated samples. As explained earlier, the measure in (3.12) has the advantage of being comparable for vectors of different sizes. In particular, suppose we are interested in the marginal contribution of information from a variable with respect to individual parameters. Let $\boldsymbol{\Omega}^{k,k}$ be the k-th diagonal element of $\boldsymbol{\Omega}^{k,k}$, i.e. the CRLB for θ_k. Then, the efficiency gain with respect to θ_k from observing $\boldsymbol{y}^i$ is given by

$$\mathrm{EG}_{\theta_k}(\boldsymbol{y}^i|\bar{\boldsymbol{y}}^i) = \left(\frac{\boldsymbol{\Omega}_{\boldsymbol{\theta}}^{k,k}(\bar{\boldsymbol{y}}^i) - \boldsymbol{\Omega}_{\boldsymbol{\theta}}^{k,k}(\boldsymbol{y})}{\boldsymbol{\Omega}_{\boldsymbol{\theta}}^{k,k}(\bar{\boldsymbol{y}}^i)} \right) \times 100. \tag{3.13}$$

In the context of DSGE models, the efficiency gains measure (3.13) is used in Iskrev (2010) to assess the importance of different observed variables with respect to individual parameters. An equivalent measure, formulated in terms of efficiency loss instead of efficiency gain, is used in Palm and Nijman (1984) to assess the loss of efficiency with respect to individual parameters due to missing observations in dynamic regression models. More broadly, FIM-based criteria similar to (3.12) and (3.13) are widely used in the experiment design literature to select an optimal design, i.e. a design which maximizes, according to a given criterion, the amount of information that an experimenter can expect to learn about the parameters through an experiment (see e.g. Silvey (1980) and

Pronzato and Pázman (2013)).

Two further remarks are worth making at this point. First, the efficiency gain measure in (3.13) is only meaningful when parameter θ_k is identified, at least locally, from the full set of observables $\boldsymbol{y}$. That is, when $\boldsymbol{\Omega}_{\boldsymbol{\theta}}^{k,k}(\boldsymbol{y})$ is finite. It is possible that excluding some variable $\boldsymbol{y}^i$ from $\boldsymbol{y}$ makes θ_k unidentified so that $\boldsymbol{\Omega}_{\boldsymbol{\theta}}^{k,k}(\overline{\boldsymbol{y}}^i) = \infty$. In that case we can set $\mathrm{EG}_{\theta_k}(\boldsymbol{y}^i|\overline{\boldsymbol{y}}^i) = 100\%$, meaning that θ_k becomes identified when $\boldsymbol{y}^i$ is added to the set of observed variables. Second, the efficiency gain measure is not limited to use only with models estimated by MLE. From the so-called Bernstein-Von Mises theorem (see e.g. Walker (1969), Chen (1985), Kim (1998)), we know that Bayesian estimation procedures asymptotically inherit the properties of the classical MLE, i.e. the posterior distribution is asymptotically normal, centered at the MLE with covariance matrix equal to the inverse of the FIM. Furthermore, the asymptotic normality implies that the posterior mean and mode are asymptotically equal and converge to the MLE. Therefore, the efficiency gain measure could be evaluated at either one of these points depending on which parameter values one wishes to focus on. Of course, in practice the prior distribution does play a role, contributing information about the parameters beyond the information contained in the likelihood function. However, information in the prior is independent from that in the sample and is therefore irrelevant to the question about the relative contributions of information by different variables with respect to estimated parameters.

4 Application: The information content of asset prices

This section evaluates the contribution of information by asset price variables in a DSGE model containing news shocks. In particular, I am interested in the validity of the following two claims: (1) standard macroeconomic variables are uninformative about news shocks; and (2) asset prices contribute a lot of information about news shocks. Clearly it is not possible to give a single general answer as to whether these statements are true or not under all circumstances. Instead, the main purpose here is to demonstrate how the measures from Section 3 can be used to study the information properties of observables in the context of a particular environment. The model I consider is taken from Schmitt-Grohé and Uribe (2012) (SGU henceforth). It is a closed economy real

business cycle model augmented with real rigidities in consumption, investment, capital utilization, and wage setting. The details of the model are given in the Online Appendix. Here I only describe those of its feature what would directly relevant for the analysis which follows. Firstly, the model has seven fundamental shocks: to neutral productivity (stationary and non-stationary), investment-specific productivity (stationary and non-stationary), government spending, wage markups and preferences. Each one of the shocks is driven by three independent innovations, two anticipated and one unanticipated. More concretely, the process governing shock x_t is given by

$$\ln(x_t/x) = \rho_x \ln(x_{t-1}/x) + \sigma_x^0 \varepsilon_{x,t}^0 + \sigma_x^4 \varepsilon_{x,t-4}^4 + \sigma_x^8 \varepsilon_{x,t-8}^8, \tag{4.1}$$

where $\varepsilon_{x,t}^j$ for $j = 0, 4, 8$ are independent standard normal random variables. The anticipated innovations $\varepsilon_{x,t-4}^4$ and $\varepsilon_{x,t-8}^8$ are known to agents in periods $t-4$ and $t-8$, respectively. Thus, they are interpreted as news shocks. SGU estimate the model using US data on the growth rates of output, consumption, investment, government expenditure, the relative price of investment, total factor productivity, and hours worked. In addition to these variables, the model makes predictions about the behavior of two asset price variables: the value of the firm and the risk-free real interest rate. In estimation, the growth rate of the value of the firm can be matched to the growth rate of the real per capita value of the stock market. Similarly, data on the risk-free real interest rate can be obtained by deflating the nominal rate on the three-month Treasury bill by the inflation rate implied by the GDP deflator. The reason SGU give for not using asset price data is that models like the one sketched here are not well suited for explaining the behavior of these variables. In other words, variables like the stock price index are not a good empirical match for the theoretical concept represented by the value of the firm.[11] Here, I abstract from the issue of whether value of the firm and the risk-free real interest rate have adequate empirical counterparts. The question I ask is whether observing these variables, if such data were available, would provide a significant amount of additional information with respect to news shocks. This question is addressed next.

[11]Some authors, such as Avdjiev (2016), who use asset prices to estimate models with news shocks, deal with this discrepancy by assuming that the data are contaminated by measurement errors. I replicate the analysis in this section for the model in Avdjiev (2016) in the Online Appendix.

4.1 Information about news shocks

Following the notation in Section 3, let $\boldsymbol{y}$ be a vector collecting the observations of all observable variables (including v^f and r), and $\bar{\boldsymbol{y}}$ be the vector of observations of the variables used in the baseline estimation of SGU, i.e. $\bar{\boldsymbol{y}} = \boldsymbol{y} \setminus (v^f, r)$. Note that $\boldsymbol{y}$ is a $T \times 9$ dimensional vector, and $\bar{\boldsymbol{y}}$ is a $T \times 7$ dimensional vector. The purpose of this section is to evaluate the information gains from observing v^f, r, or both, with respect to news shocks, which in this model are represented by the anticipated innovations to the seven fundamental shocks. There are 14 such innovations, each one of which is a T dimensional vector. I set $T = 207$, which is the sample size in SGU.

SGU solve the model by log-linear approximation of the equilibrium conditions around steady state. The linearity of the solution together with the assumption that the structural innovations and the measurement error in output growth are Gaussian, implies that the joint distribution of (any subset of) the innovations, shocks, state and observed variables is also Gaussian. This fact is used in SGU to compute the likelihood function required for estimation of the model parameters with classical and Bayesian methods. In addition, it implies that the information gains measures discussed in Section 3.1 can be computed analytically for a given set of parameter values. In the analysis which follows I fix the parameter values at the MLE reported in Schmitt-Grohé and Uribe (2012) (see Table B1 in the Online Appendix). As I show in the Online Appendix, the main conclusions do not change in any significant way if the mean or the mode of the posterior distribution are used instead.

Table 1 presents the results for all innovations – anticipated and unanticipated. The first two columns show the unconditional gains from observing $\bar{\boldsymbol{y}}$, and the additional gains from observing both v^f and r, conditional on $\bar{\boldsymbol{y}}$ being observed. Note that the information gains from observing all nine variables are given by $\mathrm{IG}_\varepsilon(\boldsymbol{y}) = \mathrm{IG}_\varepsilon(\bar{\boldsymbol{y}}) + \mathrm{IG}_\varepsilon(v^f, r|\bar{\boldsymbol{y}})$. The results show that none of the innovations, anticipated or unanticipated, can be fully recovered from the observed variables, even when v^f and r are among them. The largest reduction of uncertainty is with respect to the unanticipated stationary investment-specific productivity innovations ($\varepsilon^0_{z^I}$) – by about 94%, and the unanticipated stationary neutral productivity innovations (ε^0_z) – by about 78%. In terms of anticipated innovations, i.e. news shocks, the information gains are largest with respect to the 8–quarter ahead preference shock – about 63%, and the 4–quarter ahead wage markup shock – about 58%. However, the contribution of information by asset prices with respect to these shocks is fairly modest. The largest gains due to observing v^f and r are with respect to news

Table 1: Information content of asset prices: innovations

	innovation	IG($\bar{\boldsymbol{y}}$)	IG($v^f, r\|\bar{\boldsymbol{y}}$)	IG($v^f\|\bar{\boldsymbol{y}}$)	IG($r\|\bar{\boldsymbol{y}}$)	IG(v^f)	IG(r)
$\varepsilon^0_{\mu^a}$	non-stat. investment-specific prod.	26.3	3.7	3.5	0.1	0.2	0.1
$\varepsilon^4_{\mu^a}$	non-stat. investment-specific prod. 4q	38.4	3.2	3.0	0.3	0.1	0.1
$\varepsilon^8_{\mu^a}$	non-stat. investment-specific prod. 8q	34.0	3.3	3.1	0.3	0.1	0.1
$\varepsilon^0_{\mu^x}$	non-stat. neutral prod.	26.9	19.9	14.6	4.3	18.4	2.2
$\varepsilon^4_{\mu^x}$	non-stat. neutral prod. 4q	2.1	8.6	2.3	5.0	0.6	1.4
$\varepsilon^8_{\mu^x}$	non-stat. neutral prod. 8q	2.3	8.9	2.0	5.6	0.6	1.6
$\varepsilon^0_{z^I}$	stat. investment-specific prod.	84.0	9.5	7.0	5.0	0.9	0.9
$\varepsilon^4_{z^I}$	stat. investment-specific prod. 4q	10.3	23.3	17.8	8.3	0.8	1.7
$\varepsilon^8_{z^I}$	stat. investment-specific prod. 8q	16.5	33.3	26.7	11.1	1.6	2.8
ε^0_{z}	stat. neutral prod.	71.7	6.5	2.6	3.3	42.2	8.5
ε^4_{z}	stat. neutral prod. 4q	2.6	11.9	1.5	8.9	0.8	2.3
ε^8_{z}	stat. neutral prod. 8q	2.7	12.0	1.5	9.0	0.8	2.3
ε^0_{μ}	wage markup	9.7	27.4	1.9	21.6	3.4	0.6
ε^4_{μ}	wage markup 4q	52.9	5.0	1.7	3.5	15.1	42.3
ε^8_{μ}	wage markup 8q	34.4	4.6	2.3	2.5	9.8	27.5
ε^0_{g}	government spending	28.8	2.6	0.3	2.0	0.7	0.1
ε^4_{g}	government spending 4q	47.9	1.9	0.5	1.1	0.6	1.8
ε^8_{g}	government spending 8q	18.6	0.9	0.3	0.5	0.3	0.7
ε^0_{ζ}	preference	16.3	5.3	1.6	3.2	0.6	0.1
ε^4_{ζ}	preference 4q	16.0	1.3	0.8	0.2	0.3	0.9
ε^8_{ζ}	preference 8q	60.7	2.7	1.2	0.7	1.2	3.4

Note: $\bar{\boldsymbol{y}}$ contains all observed variables ($\boldsymbol{y}$) except asset prices (v^f and r). The information gain $\mathrm{IG}_\varepsilon(\boldsymbol{x})$ measures the reduction in uncertainty about variable ε due to observing variable $\boldsymbol{x}$, in per cent of the prior (unconditional) uncertainty. The conditional information gain $\mathrm{IG}_\varepsilon(\boldsymbol{x}|\boldsymbol{z}) = \mathrm{IG}_\varepsilon(\boldsymbol{x},\boldsymbol{z}) - \mathrm{IG}_\varepsilon(\boldsymbol{z})$ measures the additional reduction in uncertainty from observing $\boldsymbol{x}$ given that $\boldsymbol{z}$ is observed.

about the stationary investment-specific productivity shocks. They are about 23% with respect to $\varepsilon^4_{z^I}$ and 33% with respect to $\varepsilon^8_{z^I}$. Other news shocks for which the contribution of asset prices is non-trivial are the stationary and non-stationary neutral productivity shocks. The gains are about 12% and 9%, respectively. The relative contributions of the two asset price variables can seen from the third and fourth columns of the table, which report the additional gains from observing either v^f or r, conditional on $\bar{\boldsymbol{y}}$ being observed. Even though both asset price variables contribute a substantial amount of information about $\varepsilon^4_{z^I}$ and $\varepsilon^8_{z^I}$, the gains from observing v^f are significantly larger. At the same time, r is the more informative of the two variables with respect to the news components of the stationary and non-stationary neutral productivity shocks. The last two columns show the information gains from observing either v^f or r, relative to having no data at all. Interestingly, the gains are very small with respect to most news shocks, including the

stationary investment-specific productivity shocks and the neutral productivity shocks. This means that almost all of the information which asset prices contribute with respect to these shocks comes from the interactions of v^f and r with variables in $\bar{\boldsymbol{y}}$. That is, the information comes from cross-moments of asset prices and macro variables rather than the own moments of asset prices. Note that the opposite is true in the case of the news components in the wage markup shock. The unconditional gains from observing v^f or r are much larger than the conditional gains. This implies that most of the information provided by either one of the asset price variables is also contained in other observed variables.[12]

The results in Table 1 raise the question of how v^f and r compare to other observables in terms of the amount of information they provide about news shocks. To answer this question, I compute conditional information gains for each variable in $\boldsymbol{y}$. That is, I evaluate $\mathrm{IG}_\varepsilon(x|\boldsymbol{y}^x)$ where x is one of the nine observables, and $\boldsymbol{y}^x$ contains all observables (including v^f and r) except x. The results are shown in Figure 1. Note that hours worked and TFP each contribute more information with respect to the stationary investment-specific productivity news shocks compared to v^f or r. The relative price of investment is by far the most informative variable with respect to the news components in the non-stationary investment-specific productivity shocks, while government expenditure and consumption are, respectively, the most informative variables about government spending and preference news shocks. The anticipated innovations to the stationary neutral productivity shocks are the only news shocks for which an asset price variable, specifically the risk-free rate, contributes significantly more information than any other variable.

An important conclusion that emerges from the results in Table 1 and Figure 1 is that the variables' contributions of information could depend on what other variables are observed. In some cases, the amount of additional information brought by a variable is increased due to presence of other variables; in other cases the contribution is diminished. For instance, as we saw in Table 1, v^f alone provides very little information about $\varepsilon^4_{z^I}$ and $\varepsilon^8_{z^I}$. Conditional on observing all seven macro variables, however, the contribution of v^f is substantial. The opposite is true with respect to $\varepsilon^0_{\mu^x}$ and ε^0_z. This suggests that there exists a degree of positive information complementarity between v^f and (some of the) macro variables in the first case, and negative complementarity, or information

[12]The unanticipated innovation to the non-stationary neutral productivity shock $\varepsilon^0_{\mu^x}$ is an example of a third possibility – where the information gains from observing v^f are relatively large, both conditionally and unconditionally.

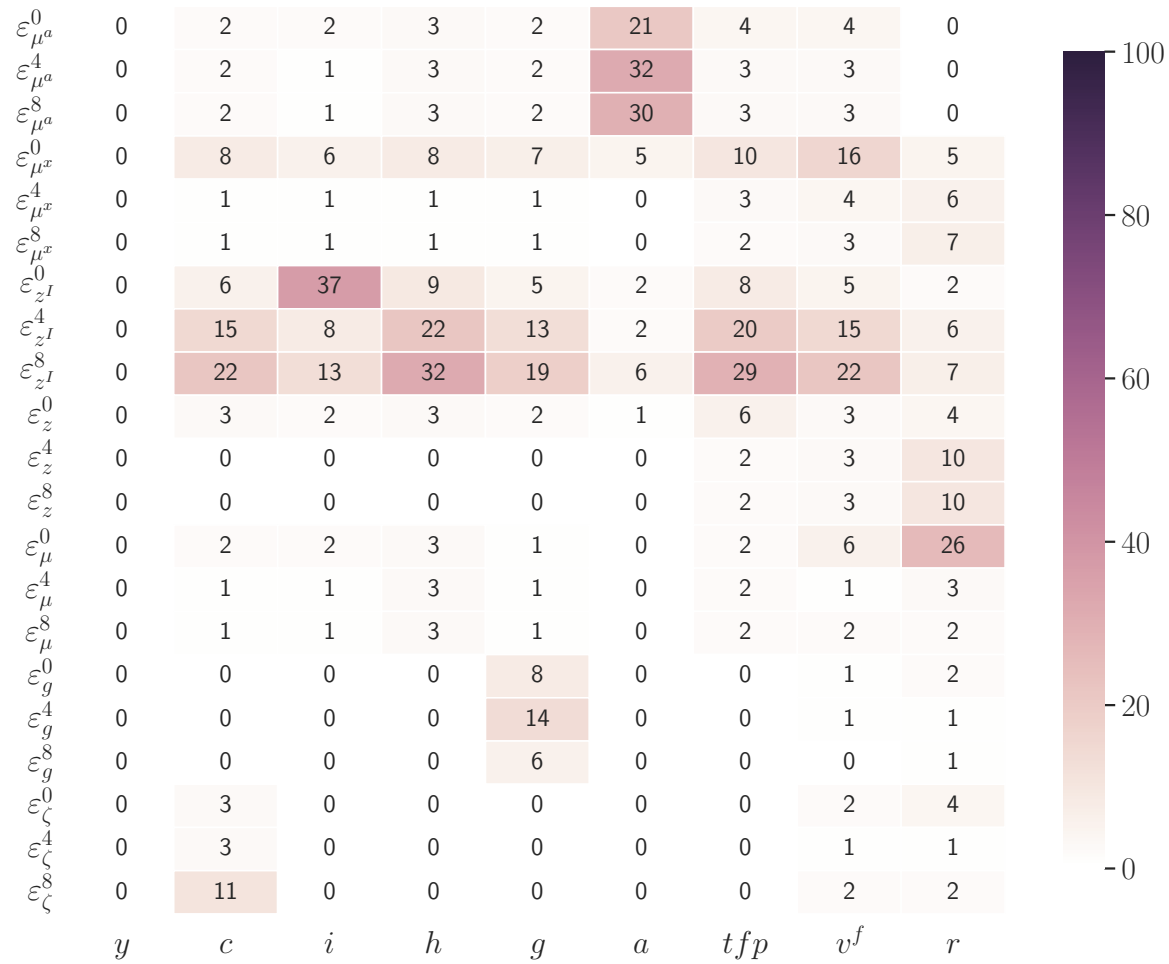

Figure 1: Conditional information gains at MLE of SGU.

redundancy, in the second. To find out how v^f interacts with each one of the macro variables, it is helpful to define a measure of (conditional) information complementarity between two variables. Specifically, let x be a member of $\bar{\boldsymbol{y}}$ and $\bar{\boldsymbol{y}}^x = \bar{\boldsymbol{y}} \setminus x$. Then, the conditional information complementarity with respect to variable ε between v^f and x can be defined as:

$$\mathrm{IC}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) = \frac{\mathrm{IG}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x)}{\mathrm{IG}_\varepsilon(v^f|\bar{\boldsymbol{y}}^x) + \mathrm{IG}_\varepsilon(x|\bar{\boldsymbol{y}}^x)} - 1. \tag{4.2}$$

Negative values indicate negative complementarity, or information redundancy, between v^f and x, and positive values indicate positive complementarity between the two vari-

ables. Since the information gain is non-negative, we have $\mathrm{IC}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) \geq -1/2$, with equality when v^f and x are (conditionally on $\bar{\boldsymbol{y}}^x$) functionally dependent, in which case $\mathrm{IG}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) = \mathrm{IG}_\varepsilon(v^f|\bar{\boldsymbol{y}}^x) = \mathrm{IG}_\varepsilon(x|\bar{\boldsymbol{y}}^x)$.[13] A lack of information complementarity, i.e. $\mathrm{IC}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) = 0$ could occur if v^f and x are (conditionally on $\bar{\boldsymbol{y}}^x$) independent, and hence $\mathrm{IG}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) = \mathrm{IG}_\varepsilon(v^f|\bar{\boldsymbol{y}}^x) + \mathrm{IG}_\varepsilon(x|\bar{\boldsymbol{y}}^x)$. Note that instead of $\bar{\boldsymbol{y}}^x$ in (4.2) the conditioning could be with respect to any other set of variables, including the empty set which would show the unconditional complementarity between v^f and x.

Using Table 1, we can determine the degree of complementarity between v^f and r, conditionally on the seven macro variables. There is a positive complementarity with respect to the preference shock, the stationary and non-stationary neutral productivity shocks, and the government spending shock. At the same time there is a negative complementarity, or redundancy of information, with respect to the stationary and non-stationary investment-specific news shocks, and wage markup shock. Overall, the degree of complementarity, both positive and negative, is relatively weak.

Figures B1 - B4 in the Online Appendix show results for conditional and unconditional information complementarity between v^f and r and each one of the macro variables. The main findings can be summarized as follows: (1) both v^f and r display very strong conditional complementarity with h and tfp, and relatively weaker, but still significant complementarity with c; (2) The complementarity is positive with respect to the news components in the stationary and non-stationary investment specific shocks, and, in the case of h and c, the stationary and non-stationary neutral productivity shocks. The complementarity is negative with respect to news about the wage markup and preference shocks; (3) The magnitude and even the sign of the information complementarity may change depending on the conditioning variables. For instance, unconditionally, v^f is strongly complementary only with tfp, and the complementarity is positive with respect to all news shocks except the two neutral productivity news shocks. r, on the other hand, is unconditionally strongly complementary primarily with h, and the complementarity is positive with respect to all news except the wage markup news shocks; (4) Conditionally, there is zero information complementarity between either v^f or r, on one hand, and y, on the other.

As noted earlier, it is not possible to recover without error the 21 anticipated and unanticipated innovations from either 7 or 9 observed variables. In fact, in many cases

[13] Note that $\mathrm{IG}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) \geq \max\left(\mathrm{IG}_\varepsilon(v^f|\bar{\boldsymbol{y}}^x), \mathrm{IG}_\varepsilon(x|\bar{\boldsymbol{y}}^x)\right)$ and $\mathrm{IG}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x) = 0$ implies $\mathrm{IG}_\varepsilon(v^f|\bar{\boldsymbol{y}}^x) = \mathrm{IG}_\varepsilon(x|\bar{\boldsymbol{y}}^x) = 0$. In that case $\frac{\mathrm{IG}_\varepsilon(v^f, x|\bar{\boldsymbol{y}}^x)}{\mathrm{IG}_\varepsilon(v^f|\bar{\boldsymbol{y}}^x) + \mathrm{IG}_\varepsilon(x|\bar{\boldsymbol{y}}^x)} = \frac{0}{0}$, which is taken to be equal to 1.

Table 2: Information content of asset prices: shocks

	shock	IG($\bar{\boldsymbol{y}}$)	IG($v^f, r|\bar{\boldsymbol{y}}$)	IG($v^f|\bar{\boldsymbol{y}}$)	IG($r|\bar{\boldsymbol{y}}$)	IG(v^f)	IG(r)
μ^a	nonstationary investment-specific prod.	100.0	0.0	0.0	0.0	0.2	0.2
μ^x	nonstationary neutral prod.	30.8	16.8	12.2	4.1	19.2	2.6
z^I	stationary investment-specific prod.	86.4	11.0	10.6	2.5	2.1	2.9
z	stationary neutral prod.	76.5	5.8	4.1	1.5	43.7	8.8
μ	wage markup	98.2	1.4	0.8	1.0	28.7	68.1
g	government spending	98.8	0.3	0.2	0.1	1.5	1.9
ζ	preference	98.3	1.3	0.7	0.5	2.2	3.7

Note: see note to Table 1

the information gains are small, meaning that the posterior uncertainty remains very close to the prior uncertainty. There are, however, only 7 structural shocks and it is natural to expect that they are easier to recover than the innovations. This is indeed the case, as can be seen in Table 2, which shows results from the same analysis as in Table 1, now applied to the structural shocks. With 9 observed variables the information gains exceed 97% for 5 of the shocks. The two shocks for which the gains are relatively small are the non-stationary neutral productivity – around 48%, and the stationary neutral productivity – around 82% with 9 observed variables. The asset price variables provide a significant amount of additional information with respect to the non-stationary neutral productivity and the stationary investment-specific productivity shocks. Most of these gains are due to information in v^f. The information gains are 100% with respect to the non-stationary investment-specific productivity, meaning that the realizations of μ_t^a can be completely recovered from the observed variables. This is a consequence of the assumption that the technology which converts consumption into investment goods is linear. As a result, in equilibrium the growth rate of the relative price of investment is equal to non-stationary investment-specific productivity shock. Hence, observing the price of investment alone is sufficient to fully recover μ_t^a for all t. None of the other shocks can be fully recovered from the observed variable, although the information gains exceed 99% in the case of wage markup, government expenditures, and preference shocks.

4.2 Information about parameters

This section evaluates the information content of asset prices with respect to the news shock–related parameters in the model. It supplements the analysis in Schmitt-Grohé and

Uribe (2012) who show that the parameters are identified from the second-order moments of the seven variables used in estimation.[14] It is clear that having additional observed variables would increase the amount of information. The purpose of the following analysis is to provide a quantitative assessment of the size of the gains from observing v^f, r, or both, and to compare them to the information gains from other observables.

As discussed in Section 3.2, the information content of a variable (or a set of variables) x with respect to estimated parameters is measured in terms of efficiency gains, which are computed using the parameter CRLBs with and without x. The differences between the values of the bounds reflect the information content of the model-implied restrictions on the joint distribution of x and the other observables. Hence, parameters for which these restrictions are more informative will see a greater reduction in the values of their lower bounds, i.e. larger efficiency gains.

The results are presented in Table 3.[15] As also discussed in Section 3.2, the gains are in terms of reduction in uncertainty as a per cent of the uncertainty conditional on observing $\bar{\boldsymbol{y}}$. Overall, the efficiency gains are substantial, in the order of between 90% and 97% for parameters of news shocks when both v^f and r are included. The gains are smaller but still significant when only one of the asset price variables is observed, and tend to be larger if that variable is v^f. These results seem to suggest that asset prices are indeed very informative with respect to news shock–related parameters. However, this does not imply that v^f and r are more informative than other observables. To find out if they are, one has to compare the efficiency gains from observing asset prices to the gains from other variables. Figure 2 does that for each one of the nine variables in $\boldsymbol{y}$. Note that, unlike in Table 3, the efficiency gains are now relative to eight, not seven, observables. For instance, the gains from observing r are relative to observing all other variables, including v^f. As a result, they are generally much smaller than before, especially with respect to news shocks parameters. This means that once v^f is observed, r adds relatively little new information about these parameters. At the same time, the efficiency gains due to v^f remain substantial, although not as large as in Table 3. Comparing the results across all variables shows that h and tfp tend to be as informative with respect to news shocks–related parameters as v^f and r. Only in the cases of the neutral productivity shocks – both stationary and non-stationary, are the asset price

[14] Even though Schmitt-Grohé and Uribe (2012) de-mean the data, this does not result in loss of information since all parameters for which first-order moments are informative are assumed to be known, i.e. are calibrated and not estimated.

[15] The results for all parameters can be found in Table B3 of the Online Appendix.

Table 3: Efficiency gains (%)

	parameter	v^f, r	v^f	r
σ_z^0	std. stationary neutral productivity	87	59	83
σ_z^4	std. stationary neutral productivity q4	93	73	72
σ_z^8	std. stationary neutral productivity q8	90	73	64
$\sigma_{\mu^a}^0$	std. non-stationary investment-specific productivity	95	95	43
$\sigma_{\mu^a}^4$	std. non-stationary investment-specific productivity q4	97	96	74
$\sigma_{\mu^a}^8$	std. non-stationary investment-specific productivity q8	96	96	68
σ_g^0	std. government spending	97	80	95
σ_g^4	std. government spending q4	91	89	52
σ_g^8	std. government spending q8	91	89	55
$\sigma_{\mu^x}^0$	std. non-stationary neutral productivity	78	50	67
$\sigma_{\mu^x}^4$	std. non-stationary neutral productivity q4	94	78	71
$\sigma_{\mu^x}^8$	std. non-stationary neutral productivity q8	90	74	58
σ_μ^0	std. wage markup	99	70	97
σ_μ^4	std. wage markup q4	90	84	52
σ_μ^8	std. wage markup q8	90	85	48
σ_ζ^0	std. preference	98	89	97
σ_ζ^4	std. preference q4	90	88	39
σ_ζ^8	std. preference q8	91	88	50
$\sigma_{z^I}^0$	std. stationary investment-specific productivity	92	91	66
$\sigma_{z^I}^4$	std. stationary investment-specific productivity q4	96	93	81
$\sigma_{z^I}^8$	std. stationary investment-specific productivity q8	92	88	72

Note: The efficiency gain $\mathrm{EG}_{\theta_i}(\boldsymbol{x}|\bar{\boldsymbol{y}})$, for (1) $\boldsymbol{x} = (v^f, r)$, (2) $\boldsymbol{x} = v^f$, or (3) $\boldsymbol{x} = r$, is defined as the reduction in the value of CRLB for θ_i when all variables are observed, as a per cent of the value of the CRLB when all variables except those in $\boldsymbol{x}$ are observed.

variables the most informative ones, with tfp being close next best. Therefore, we can conclude that, although very informative about most model parameters, the two asset price variables in this model are not in any way uniquely important for the identification of news shocks–related parameters.

Redundancy of output. An interesting result that emerged from the above analysis is that output growth data does not contribute any additional information with respect to either latent variables (i.e. innovations and shocks) or the free parameters in the model. In other words y is redundant given the other observed variables. In fact, it can be shown that, as long as the growth rates of consumption, investment and government expenditures are observed, the output growth variable is only informative with respect to one parameter – the standard deviation of the measurement error $\sigma_{g^y}^{me}$. Two assumptions in SGU are responsible for this result: (1) several model parameters are known, and (2) output is the only variable observed with measurement error. Relaxing either one of these assumptions would make output growth informative. For more details, see Iskrev (2015).

To summarize, the analysis in this section shows that the two asset price variables in the SGU model are not particularly informative about either the realizations of news shocks or parameters related to news shocks. Macroeconomic aggregates, such as hours worked, TFP, or the relative price of investment, are about as informative with respect to news shocks as are the value of the firm (stock prices) or the risk-free interest rate. Needless to say, this is a conclusion about the properties of the estimated SGU model. Making changes that affect the way news shocks propagate throughout the economy could alter the results. For instance, different values of the model parameters could imply a much larger information content of asset prices. I explore this possibility in the Online Appendix, by considering alternative parameterizations of the SGU model, taken from Herbst and Schorfheide (2014), who estimate the same model with the same set of variables using a different estimation approach, and from Miyamoto and Nguyen (2015), who estimate the model adding forecast data to the original set of macroeconomic variables. The results show that the information gains with respect to news shocks due to observing asset prices remain small. Instead of changing the parameter values, one may modify the way news shocks are introduced into the model. While in most of the existing literature news shocks are specified similarly to the SGU model, as anticipated innovations to fundamental shocks, Avdjiev (2016) considers an alternative specification

where news are interpreted as anticipated changes in the long-run levels of fundamental shocks. He finds that the long-run specification has stronger empirical support than the alternative. Importantly, Avdjiev (2016) estimates his model using asset price data, namely, the growth rate of the total stock market valuation and the real risk-free rate. To find out if these changes alter the conclusion about the role of asset prices, in the Online Appendix I apply the analysis of this section to the model of Avdjiev (2016). In summary, the results show that the informational importance of asset prices in the Avdjiev (2016) model depends crucially on whether or not TFP is treated as observed. Assuming that TFP is unobserved, as Avdjiev (2016) does, results in large contributions of information by both asset price variables. On the other hand, if TFP is treated as observed, as in SGU, only the interest rate is found to be more informative than any other observable with respect to a single new shock, namely, the stationary neutral productivity news shock. This can be explained with the fact that there exists significant complementarity between the asset price variables, on one hand, and TFP, on the other. For more details see Section C in the Online Appendix.

5 Conclusion

The informational importance of observed variables with respect to structural shocks, or unobserved endogenous variables, in business cycle models is often asserted without a formal justification. This paper has proposed a general framework for measuring the contribution of information that different observed variables make with respect to a given latent variable. This allows researchers to evaluate and compare the informational value of observables and identify the most informative ones. Having well identified structural shocks and key unobserved endogenous variables, like potential output or natural rate of interest, is a critical requirement for DSGE models to meet in order to fulfill their potential as credible story-telling devices. Thus, the methodology described in the paper could benefit both researchers who develop and estimate DSGE models and the readers of such research, by improving their understanding and increasing the transparency of these models.

An application to a business cycle model featuring news shocks revealed a relatively modest contribution of information by asset prices. A necessary caveat to this result is that it is entirely conditional on the particular model considered. Making changes in the way shocks are introduced and propagate, or in the way asset prices are modeled, is likely

to have an impact on the conclusions regarding the informational value of observables. Indeed, this is an example of one of the intended purposes of the analysis developed in this paper, namely, checking whether models are consistent with our intuition about how the real world works. Finding out that they are not provides useful directions for their improvement.

The analysis in this paper can be extended in several directions. With regards to news-driven DSGE models, it would be interesting to know whether observing expectations provides significantly more information than observing asset prices. Some evidence that the identification of news shocks improves when data on expectations are used is provided in Miyamoto and Nguyen (2019). However, their analysis only demonstrates that information increases when relevant variables are added, and not that expectations are superior sources of information about news shocks than other observables.[16] In terms of the methodology itself, more research is needed on how to perform this type of analysis in non-linear and non-Gaussian models. In particular, the information gains measures are, in general, not available in closed form, and would have to be estimated using simulated data.

[16] Miyamoto and Nguyen (2019) find that the posterior distributions of the contributions of news shocks to the unconditional variances of macro aggregates are tighter when data on expectations are used, compared to when the same model is estimated without expectations. In principle, this only shows that the additional variables are non-redundant with respect to the estimated parameters. The methodology described in this paper can be used to show how the informational contributions of expectations compare to other observables.

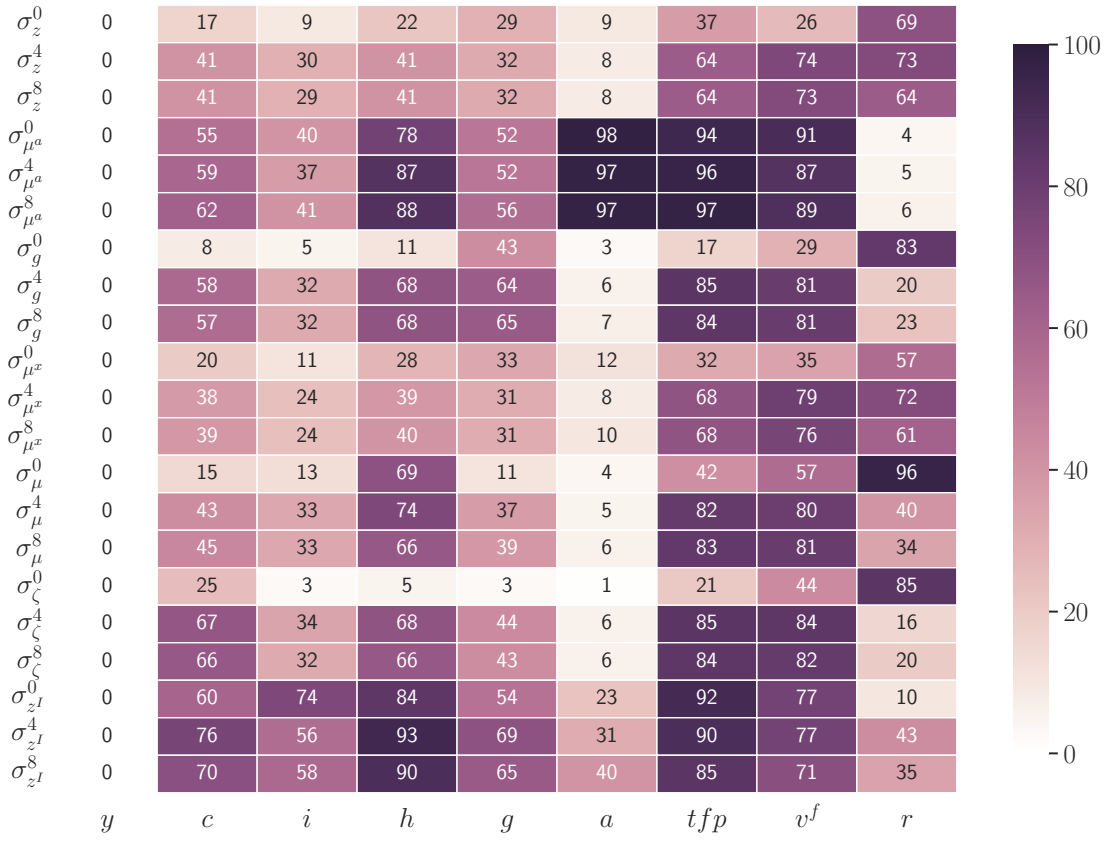

Figure 2: Efficiency gains at MLE of SGU.

References

Ajello, A. (2016): "Financial intermediation, investment dynamics, and business cycle fluctuations," *American Economic Review*, 106, 2256–2303.

Alessi, L., M. Barigozzi, and M. Capasso (2011): "Non-Fundamentalness in Structural Econometric Models: A Review," *International Statistical Review*, 79, 16–47.

Amblard, P.-O. and O. J. Michel (2011): "On directed information theory and Granger causality graphs," *Journal of computational neuroscience*, 30, 7–16.

Avdjiev, S. (2016): "News Driven Business Cycles and data on asset prices in estimated DSGE models," *Review of Economic Dynamics*, 20, 181–197.

Barnett, L., A. B. Barrett, and A. K. Seth (2009): "Granger causality and transfer entropy are equivalent for Gaussian variables," *Physical review letters*, 103, 238701.

Beaudry, P., P. Fève, A. Guay, and F. Portier (2015): "When is Nonfundamentalness in VARs a Real Problem? An Application to News Shocks," Tech. rep., National Bureau of Economic Research.

Beaudry, P. and F. Portier (2006): "Stock Prices, News, and Economic Fluctuations," *The American Economic Review*, 96, pp. 1293–1307.

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

Bretscher, L., A. Malkhozov, and A. Tamoni (2019): "News Shocks and Asset Prices," *Available at SSRN 2367196*.

Chahrour, R. and K. Jurado (2017): "Recoverability," Tech. rep.

Chen, C.-F. (1985): "On asymptotic normality of limiting density functions with Bayesian implications," *Journal of the Royal Statistical Society. Series B (Methodological)*, 540–546.

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

Cover, T. M. and J. A. Thomas (2006): *Elements of information theory, 2nd edition*, John Wiley & Sons.

Davis, J. (2007): "News and the term structure in general equilibrium," Working papers, Northwestern University.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B*, 1–38.

Fernandez-Villaverde, J., J. Rubio-Ramirez, T. J. Sargent, and M. W. Watson (2007): "A, B, Cs (and Ds) of Understanding VARs," *American Economic Review*, 97.

Forni, M. and L. Gambetti (2014): "Sufficient information in structural VARs," *Journal of Monetary Economics*, 66, 124–136.

Forni, M., L. Gambetti, and L. Sala (2016): "VAR Information and the Empirical Validation of DSGE Models," *Centre for Economic Policy Research*, discussion Paper 11178.

Franchi, M. and P. Paruolo (2015): "Minimality of state space solutions of DSGE models and existence conditions for their VAR representation," *Computational Economics*, 46, 613–626.

Franchi, M. and A. Vidotto (2013): "A check for finite order VAR representations of DSGE models," *Economics Letters*, 120, 100–103.

Fujiwara, I., Y. Hirose, and M. Shintani (2011): "Can News Be a Major Source of Aggregate Fluctuations? A Bayesian DSGE Approach," *Journal of Money, Credit and Banking*, 43, 1–29.

Geweke, J. (1982): "Measurement of linear dependence and feedback between multiple time series," *Journal of the American statistical association*, 77, 304–313.

Geweke, J. F. (1984): "Measures of conditional linear dependence and feedback between time series," *Journal of the American Statistical Association*, 79, 907–915.

Giacomini, R. (2013): "The relationship between DSGE and VAR models," *Advances in Econometrics*, 31.

Giannone, D. and L. Reichlin (2006): "Does information help recovering structural shocks from past observations?" *Journal of the European Economic Association*, 4, 455–465.

Gilchrist, S., A. Ortiz, and E. Zakrajsek (2009a): "Credit risk and the macroeconomy: Evidence from an estimated dsge model," *Unpublished manuscript, Boston University*, 13.

Gilchrist, S., V. Yankov, and E. Zakrajšek (2009b): "Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets," *Journal of monetary Economics*, 56, 471–493.

Gilchrist, S. and E. Zakrajšek (2012): "Credit spreads and business cycle fluctuations," *American Economic Review*, 102, 1692–1720.

Görtz, C. and J. D. Tsoukalas (2017): "News and financial intermediation in aggregate fluctuations," *Review of Economics and Statistics*, 99, 514–530.

Granger, C. and J.-L. Lin (1994): "Using the mutual information coefficient to identify lags in nonlinear models," *Journal of time series analysis*, 15, 371–384.

Granger, C. W. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, 424–438.

Hansen, H. and T. Sargent (1991): "Two Difficulties in Interpreting Vector Autoregressions," in *Rational Expectations Econometrics*, ed. by H. Hansen and T. Sargent, Westview Press.

Hansen, L. P. and T. J. Sargent (1980): "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and control*, 2, 7–46.

Herbst, E. and F. Schorfheide (2014): "Sequential Monte Carlo sampling for DSGE models," *Journal of Applied Econometrics*, 29, 1073–1098.

Hirose, Y. and T. Kurozumi (2012): "Identifying News Shocks with Forecast Data," CAMA Working Papers 2012-01, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Ilut, C. L. and M. Schneider (2014): "Ambiguous business cycles," *American Economic Review*, 104, 2368–99.

Iskrev, N. (2010): "Evaluating the strength of identification in DSGE models. An a priori approach," Working paper series, Banco de Portugal.

——— (2015): "What's News in Business Cycles? A Comment," Tech. rep.

Jarociński, M. and B. Maćkowiak (2017): "Granger causal priority and choice of variables in vector autoregressions," *Review of Economics and Statistics*, 99, 319–329.

Joe, H. (1989): "Relative entropy measures of multivariate dependence," *Journal of the American Statistical Association*, 84, 157–164.

Kaihatsu, S. and T. Kurozumi (2014): "Sources of business fluctuations: Financial or technology shocks?" *Review of Economic Dynamics*, 17, 224–242.

Kim, J.-Y. (1998): "Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models," *Econometrica*, 359–380.

Kruskal, W. (1987): "Relative Importance by Averaging Over Orderings," *The American Statistician*, 41, 6–10.

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

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013): "Fiscal foresight and information flows," *Econometrica*, 81, 1115–1145.

Lippi, M. and L. Reichlin (1993): "The dynamic effects of aggregate demand and supply disturbances: Comment," *The American Economic Review*, 83, 644–652.

——— (1994): "VAR analysis, nonfundamental representations, Blaschke matrices," *Journal of Econometrics*, 63, 307–325.

Liu, Z., P. Wang, and T. Zha (2013): "Land-price dynamics and macroeconomic fluctuations," *Econometrica*, 81, 1147–1184.

Meng, X.-L. and X. Xie (2014): "I got more data, my model is more refined, but my estimator is getting worse! Am I just dumb?" *Econometric Reviews*, 33, 218–250.

Milani, F. and A. Rajbhandari (2012): "Observed expectations, news shocks, and the business cycle," *University of California–Irvine Department of Economics Working Paper*, 121305.

Milani, F. and J. Treadwell (2012): "The effects of monetary policy "news" and "surprises"," *Journal of Money, Credit and Banking*, 44, 1667–1692.

Miyamoto, W. and T. L. Nguyen (2015): "News shocks and Business cycles: Evidence from forecast data," Tech. rep.

——— (2019): "The Expectational Effects of News in Business Cycles: Evidence from Forecast Data," *Journal of Monetary Economics, forthcoming*.

Palm, F. C. and T. E. Nijman (1984): "Missing observations in the dynamic regression model," *Econometrica: journal of the Econometric Society*, 1415–1435.

Peña, D. and A. v. d. Linde (2007): "Dimensionless measures of variability and dependence for multivariate continuous distributions," *Communications in Statistics - Theory and Methods*, 36, 1845–1854.

Peña, D. and J. Rodríguez (2003): "Descriptive measures of multivariate scatter and linear dependence," *Journal of Multivariate Analysis*, 85, 361 – 374.

Pourahmadi, M. and E. S. Soofi (2000): "Prediction variance and information worth of observations in time series," *Journal of Time Series Analysis*, 21, 413–434.

Pronzato, L. and A. Pázman (2013): "Design of experiments in nonlinear models," *Lecture Notes in Statistics*, 212.

Rao, C. R. (2002): *Linear statistical inference and its applications*, John Wiley & Sons.

Ravenna, F. (2007): "Vector autoregressions and reduced form representations of DSGE models," *Journal of monetary economics*, 54, 2048–2064.

Retzer, J. J., E. S. Soofi, and R. Soyer (2009): "Information importance of predictors: Concept, measures, Bayesian inference, and applications," *Computational Statistics & Data Analysis*, 53, 2363–2377.

Schmitt-Grohé, S. and M. Uribe (2012): "What's News in Business Cycles," *Econometrica*, 80, 2733–2764.

Schreiber, T. (2000): "Measuring information transfer," *Physical review letters*, 85, 461.

Shannon, C. E. (1948): "A mathematical theory of communication," *Bell System Technical Journal*, 27, 379–423.

Silvey, S. (1980): *Optimal design: an introduction to the theory for parameter estimation*, Chapman and Hall, London.

Sims, E. R. (2012): "News, non-invertibility, and structural VARs," *Advances in Econometrics*, 28, 81.

Soccorsi, S. (2016): "Measuring nonfundamentalness for structural VARs," *Journal of Economic Dynamics and Control*, 71, 86–101.

Soofi, E. S. (1992): "A generalizable formulation of conditional logit with diagnostics," *Journal of the American Statistical Association*, 87, 812–816.

Theil, H. (1987): "How many bits of information does an independent variable yield in a multiple regression?" *Statistics & probability letters*, 6, 107–108.

Theil, H. and C. Chung (1988): "Information-theoretic measures of fit for univariate and multivariate linear regressions," *The American Statistician*, 42, 249–252.

Walker, A. (1969): "On the asymptotic behaviour of posterior distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 80–88.

Wei, W. W. (1978a): "The effect of temporal aggregation on parameter estimation in distributed lag model," *Journal of Econometrics*, 8, 237–246.

——— (1978b): "Some consequences of temporal aggregation in seasonal time series models," in *Seasonal analysis of economic time series*, NBER, 433–448.

Wilks, S. S. (1932): "Certain generalizations in the analysis of variance," *Biometrika*, 471–494.