December 8, 2021

Economic Theories and Macroeconomic Reality*

Francesca Loria[a], Christian Matthes[b,*], Mu-Chun Wang[c]

[a] *Federal Reserve Board*
[b] *Indiana University*
[c] *Deutsche Bundesbank*

Abstract

Economic theories are often encoded in equilibrium models that cannot be directly estimated because they lack features that, while inessential to the theoretical mechanism that is central to the specific theory, would be essential to fit the data well. We propose an econometric approach that confronts such theories with data through the lens of a time series model that is a good description of *macroeconomic reality*. Our approach explicitly acknowledges misspecification as well as measurement error. We highlight in two applications that household heterogeneity greatly helps to fit aggregate data, independently of whether or not nominal rigidities are considered.

JEL classifications: C32, C50, E30

Keywords: Bayesian Inference, Misspecification, Heterogeneity, VAR, DSGE

* We thank the editor Yuriy Gorodnichenko and associate editor Boragan Aruoba, an anonymous referee, Gianni Amisano, Christiane Baumeister, Joshua Chan, Siddhartha Chib, Thorsten Drautzburg, Thomas Drechsel, Ed Herbst, Dimitris Korobilis, Elmar Mertens, Josefine Quast, Frank Schorfheide, and Mark Watson as well as seminar and conference participants at the NBER-NSF Seminar on Bayesian Inference in Statistics and Econometrics, Indiana University, the National University of Singapore, the Drautzburg-Nason workshop, CEF, IAAE, and the NBER Summer Institute for their useful comments and suggestions on earlier versions of the paper. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Board, the Federal Reserve System, the Deutsche Bundesbank, the Eurosystem, or their staff.

* Corresponding author, E-mail address: matthesc@iu.edu

"Essentially, all models are wrong, but some are useful."

— George Box (Box and Draper, 1987)

"I will take the attitude that a piece of theory has only intellectual interest until it has been validated in some way against alternative theories and using actual data."

— Clive Granger (Granger, 1999)

1. Introduction

Economists often face four intertwined tasks: (i) discriminate between competing theories that are either not specified to a degree that they could be easily taken to the data in the sense that the likelihood function implied by those theories/models would be severely misspecified, or it would be too costly to evaluate the likelihood function often enough to perform likelihood-based inference, (ii) find priors for multivariate time series models that help to better forecast macroeconomic time series, (iii) use these multivariate time series models to infer the effects of structural shocks, and (iv) estimate unobserved cyclical components of macroeconomic aggregates. We introduce a method that tackles issue (i), and provides tools to help with the second, third, and fourth tasks.

To introduce and understand our approach, let us for simplicity assume we have two theories at hand. In order to use our approach, we assume that these theories, while they might generally have implications for different variables $X_{1,t}$ and $X_{2,t}$, share a common set of variables which we call X_t. Each theory implies restrictions on a Vector Autoregression (VAR, Sims, 1980) for X_t. We can figure out what these restrictions are by simulating data from each theory, which is achieved by first drawing from the prior for each theory's parameters, then simulating data from each theory, and finally estimating a VAR on X_t implied by each theory. Doing this repeatedly gives us a distribution of VAR parameters conditional on a theory - we interpret this as the prior for VAR coefficients implied by each theory.

We want to learn about these theories by studying how close a VAR estimated on data is to these theory-implied VARs. X_t might not be directly observed, but we have access to data on other variables Y_t that we assume are linked to the VAR variables X_t via an unobserved components model, in which we embed the VAR for X_t.[1] To estimate our model and learn about which theory is preferred by the data, we turn our theory-based priors for the VAR into one mixture-prior for the VAR using a set of *prior* weights. We devise a Gibbs sampler that jointly estimates all parameters in our model, the unobserved components X_t, as well as the *posterior* distribution of the weights on our mixture prior. The update from prior weights on each theory to a posterior distribution of weights tells us about the relative fit of each theory, taking into account issues such as measurement error or omitted trends that are captured by the unobserved components structure of our time series model.

Because we allow deviations of the VAR coefficients' posterior from either theory-based prior, we explicitly acknowledge misspecification. In particular, because of this feature there is no need to augment equilibrium models with ad-hoc features that are not central to the theories of interest, but could be crucial to the likelihood-based fit of an equilibrium model.

We use our approach to show that modeling heterogeneity across households substantially improves the fit of various economic theories by studying two scenarios: (i) a permanent income example where one of the theories features some agents that cannot participate in asset markets (they are "hand-to-mouth" consumers), and (ii) a New Keynesian example where one of the models features the same hand-to-mouth consumers.

In real models with heterogeneous households and aggregate shocks, a common finding is that heterogeneity does *not* matter substantially for aggregate dynamics - see for example the benchmark specification in Krusell and Smith (1998). Recent models with both heterogeneous households and nominal rigidities following Kaplan et al. (2018) can break this 'approximate aggregation' result. This still leaves the question whether *aggregate* data is better described by a heterogeneous agent model or its representative agent counterpart. To make progress on this issue, we introduce the aforementioned hand-to-mouth consumer into two standard economic theories. For the New Keynesian model, we build on Debortoli and Galí (2018), who show that

[1] This assumption captures issues such as measurement error and the situation where the actual data Y_t has a clear trend, while the variables in the theories X_t do not.

a two agent New Keynesian model can already approximate a richer heterogeneous model in terms of many aggregate implications.[2] In both applications we find that even a model with this relatively stark form of heterogeneity is substantially preferred by the data to its representative agent counterpart. This finding does thus not hinge on nominal rigidities: Even in our permanent income example, the model with heterogeneous agents is preferred, albeit by a smaller margin (the posterior mean of the weight on the two-agent version of that model is 0.65, whereas the posterior mean of the weight on the two agent version of the New Keynesian model varies from 0.88 to 1 depending on the exact specification).

The idea to generate priors from equilibrium for statistical models such as VARs is not new. Ingram and Whiteman (1994) generate a prior based on a real business cycle model for a VAR. Del Negro and Schorfheide (2004) go further by showing one can back out the posterior distribution of the parameters of interest for their model from the VAR posterior.[3] Our interest is different from theirs: We want to distinguish between possibly non-nested models that could not or should not be taken to the data directly.

By focusing on the difference between theory-implied VAR coefficients and VAR coefficients estimated from data, our approach is philosophically similar to indirect inference, where a theory-based model is estimated by bringing its implied coefficients for a statistical model as close as possible to the coefficients of the same statistical model estimated on real data. In fact, the initial application of indirect inference used VARs as statistical models (Smith, 1993).

The interpretation of an economic model as a device to generate priors for a statistical model is in line with the 'minimal econometric interpretation' of an equilibrium model that was put forth by Geweke (2010) (see also DeJong et al., 1996). We think of our models as narrative devices that can speak to the population moments encoded in VAR parameters. What sets us apart from Geweke (2010) is that we explicitly model mis-measurement and, at the same time, want

[2] Other papers that develop similarly tractable heterogeneous agent New Keynesian models include Bilbiie (2018) and the many references therein.

[3] Similar ideas have been used in a microeconometric context by Fessler and Kasy (2019). Filippeli et al. (2020) also build a VAR prior based on an equilibrium model, but move away from the conjugate prior structure for the VAR used by Del Negro and Schorfheide (2004).

to infer the best estimate of the true economic variable using information from all equilibrium models.

Our approach is generally related to the literature on inference in misspecified equilibrium models. Ireland (2004) adds statistical (non micro-founded) auto-correlated noise to the dynamics implied by his equilibrium model to better fit the data. To assess misspecification and improve upon existing economic models, Den Haan and Drechsel (2021) propose to add additional non-micro-founded shocks, while Inoue et al. (2020) add wedges in agents' optimization problems (which generally improve the fit of macroeconomic models, as described in Chari et al., 2007). Canova and Matthes (2021) use a composite likelihood approach to jointly update parameter estimates and model weights for a set of equilibrium models. We share with all those papers the general view that equilibrium models are misspecified, but our goal is not to improve estimates of parameters within an equilibrium model (like the composite likelihood approach) or to directly build a better theory (like papers in the 'wedge' literature), but rather to provide a framework which allows researchers to distinguish between various economic models and to construct statistical models informed by (combinations of) these theories. Our approach is also related to earlier work on assessing calibrated dynamic equilibrium models, such as Canova (1994), Watson (1993), and Gregory and Smith (1991).

Another contribution of our paper is to propose a tractable new class of mixture priors for VARs. As such, our work is related to recent advances in prior choice for VARs such as Villani (2009) and Giannone et al. (2019). In particular, in our benchmark specification, we exploit the conjugate prior structure introduced in Chan (2021) and extend it to our mixture setting. This structure has the advantage that it can handle large VARs well. Mixture priors have been used to robustify inference by, for example, Chib and Tiwari (1991) and Cogley and Startz (2019).

To incorporate a small number of specific features from an economic theory in a prior, one can adaptively change the prior along the lines presented in Chib and Ergashev (2009). In our context, we want to impose a broad set of restrictions from a number of theories instead. The mixture priors we introduce could also incorporate standard statistical priors for VARs such as the Minnesota prior as a mixture component (similar in spirit to the exercise in Schorfheide, 2000), as well as several "models" that only incorporate some aspects of theories, interpreted as restrictions on structural VAR parameters as in Baumeister and Hamilton (2019).

Finally, our framework is constructed to explicitly acknowledge limitations of *both* data and theories: We allow for various measurements of the same economic concepts to jointly inform inference about economic theories and we do not ask theories to explain low-frequency features of the data they were not constructed to explain. We accomplish the second feature by estimating a VAR for deviations from trends, where we use a purely statistical model for trends, which we jointly estimate with all other parameters, borrowing insights from Canova (2014).

Our work is also related to opinion pools (Geweke and Amisano, 2011), where the goal is to find weights on aggregate predictions of various possibly misspecified densities (see also Amisano and Geweke, 2017, Del Negro et al., 2016 and Waggoner and Zha, 2012).

2. Econometric Framework

Our methodology aims to incorporate prior information on relationships in the data from multiple models and, at the same time, to discriminate between these models. In practical terms, we build a mixture prior for a VAR of unobserved state variables (e.g., model-based measures of inflation, output gap, etc.), where each mixture component is informed by one specific economic theory (or equilibrium model). This VAR is embedded in a state space model to account for measurement issues. We first give a broad overview of our approach (i.e. the ingredients of our state space model) before going into the details of each step. A bird's eye illustration of our approach can be found in Appendix A.

Economic Theories

Consider a scenario where a researcher has K theories or economic models at hand that could potentially be useful to explain aspects of observed economic data. However, the theories are not necessarily specified in such a way that we can compute the likelihood function and achieve a reasonable fit of the data. This could be because, for example, the exogenous processes are not flexible enough to capture certain aspects of the data, or there are no relatively ad-hoc features such as investment adjustment costs in the models. While these features could be added to a specific model, we think it is useful to provide a framework that can test these simpler equilibrium models directly on the data. Our framework is also useful in situations where nonlinearities in models are important and the evaluation of the likelihood function is not computationally feasible.

What do we mean by a model? We follow the standard protocol in Bayesian statistics and call model i the collection of a prior distribution $f_i(\theta_i)$ for the vector of deep parameters θ_i of model i and the associated distribution of the data conditional on a specific parameter vector $f_i(X_i^t|\theta_i)$ (Gelman et al., 2013).[4] For each of the K models, we only require that we can simulate from $f_i(\theta_i)$ and $f_i(X_i^t|\theta_i)$. In Section 3.2 we show how this idea can be easily extended to situations where a researcher might only want to impose certain implications of a given theory. The simulation from the models constitutes one block or module of our algorithm.

Model-Based Priors for VARs

Given simulations from each model, we construct a prior from a conjugate family for a VAR on a set of variables common across models, that is $X^t \in X_{i,t}$, $\forall i$. The specific form of the prior for each mixture component is dictated by the practical necessity of using *natural conjugate priors* for our VAR, for which the marginal likelihood is known in closed form.[5] This mapping from simulations to VAR priors is the second block or module of our approach.

Macroeconomic Reality: Mixture Priors

We exploit a well-known result from Bayesian statistics (Lee, 2012): If the prior for a model is a mixture of conjugate priors, then the corresponding posterior is a mixture of the conjugate posterior coming from each of the priors. The weights of the mixture will be updated according

[4] A superscript denotes the history of a variable up to the period indicated in the superscript.

[5] As will become clear in this section, in theory we could use non-conjugate priors for each mixture, but then we would need to compute (conditional) marginal likelihoods for each parameter draw, a task that is infeasible in practice due to the computational cost it would come with.

to the fit (marginal likelihood) of the VAR model with each mixture component as prior. To make our approach operational, we note that the marginal likelihood for all conjugate priors commonly used for VARs is known *in closed form* (Giannone et al., 2015; Chan, 2021). This is important because in the final module of our approach we embed our VAR into a Gibbs sampler for a state space model because we want to allow for multiple measurements for each economic variable (e.g. CPI and PCE-based measures of inflation).

We first establish that indeed a mixture prior consisting of conjugate priors results in a mixture posterior of the conjugate posteriors. To economize on notation, we consider two theories here, but the extension to $K > 2$ theories is straightforward. We denote the VAR parameter vector by γ. Note that even though each equilibrium model has a unique parameter vector, there is only one parameter vector for the VAR. Our approach constructs a prior for this VAR parameter vector that encodes various economic theories.

Before going into detail, it will be useful to explicitly state the definition of a *natural conjugate prior* (Gelman et al., 2013).

Definition 2.1 (Natural Conjugate Prior)**.** *Consider a class $\mathscr{F}$ of sampling distributions $p(y|\gamma)$ and $\mathscr{P}$ a class of prior distributions. then the class $\mathscr{P}$ is* conjugate *for $\mathscr{F}$ if*

$$p(\gamma|y) \in \mathscr{P} \textit{ for all } p(.|\gamma) \in \mathscr{F} \textit{ and } p(.) \in \mathscr{P}$$

A class $\mathscr{P}$ is natural conjugate *if it is conjugate for a class of sampling distributions and has the same functional form as the likelihood function.*

Turning now to mixture priors, we start by defining the **prior**:

$$p(\gamma) = w_1 p_1(\gamma) + w_2 p_2(\gamma),$$

where $p_1(\gamma)$ and $p_2(\gamma)$ are both *conjugate* priors with prior mixture weights w_1 and $w_2 = 1 - w_1$. We denote a generic vector of data by y.

The **posterior** with a two-component mixture prior is given by

$$p(\gamma|y) = \frac{p(y|\gamma)p(\gamma)}{p(y)} = \frac{p(y|\gamma)(w_1 p_1(\gamma) + w_2 p_2(\gamma))}{p(y)},$$

where $p(y|\gamma)w_1 p_1(\gamma)$ can be rewritten as

$$p(y|\gamma)w_1 p_1(\gamma) = w_1 \underbrace{p_1(\gamma|y)}_{\propto p(y|\gamma)p_1(\gamma)} \underbrace{\int p(y|\tilde{\gamma})p_1(\tilde{\gamma})d\tilde{\gamma}}_{\equiv ML_1},$$

and where ML_1 is the marginal likelihood if one were to estimate the VAR with prior $p_1(\gamma)$ only. A corresponding equation holds for the second mixture component.

The **posterior** is thus a weighted average of the posteriors that we would obtain if we used each conjugate prior individually:[6]

$$\begin{aligned} p(\gamma|y) &= w'_1 p_1(\gamma|y) + w'_2 p_2(\gamma|y), && (1)\\ \text{where} \quad w'_i &= \frac{w_i ML_i}{\sum_{j=1}^{2} w_j ML_j}, \ \forall i = 1,2 && (2)\\ ML_i &= \int f_i(\gamma)p(y|\gamma)d\gamma && (3) \end{aligned}$$

Note that the expression shows how to easily construct draws from the mixture posterior: With probability w'_1 draw from $p_1(\gamma|y)$ and with the complementary probability draw from $p_2(\gamma|y)$. Draws from both $p_1(\gamma|y)$ and $p_2(\gamma|y)$ are easily obtained because they are conjugate posteriors. We embed this idea in a Gibbs sampler where the VAR governs the dynamics of an unobserved vector of state variables. Hence the model probabilities will vary from draw to draw. We next discuss each module of our approach in more detail, moving backwards from the last to the first step of our procedure.

[6] The equivalence between mixture priors and posteriors weighted by posterior model probabilities also appears in Cogley and Startz (2019).

2.1. *Our Time Series Model* - Macroeconomic Reality

With this general result in hand, we turn to our specific time series model. We embed the VAR described above in a more general time series model for two reasons. First, the mapping from variables in an economic model to the actually observed data is not unique - we usually have multiple measurements for the same economic concept. Should our models match inflation based on the CPI, PCE, or GDP deflator? Should the short-term interest rate in New Keynesian models be the Federal Funds rate or the three-month treasury bill rate? Should we use expenditure or income-based measures of real output (Aruoba et al., 2016)? To circumvent these issues, we treat model-based variables X_t as latent variables for which we observe various indicators Y_t. Second, economic theories might not be meant to describe the full evolution of macroeconomic aggregates, but rather only certain aspects. While this is generally hard to incorporate in statistical analyses, there is one specific aspect of macroeconomic theories that we can incorporate, namely that many theories are only meant to describe the *business cycle* and not low frequency movement.[7] We thus follow Canova (2014) and allow for unobserved components that are persistent and not related to the economic theories we consider.[8] Our time series model is thus:

$$Y_t = \mu + AX_t + Bz_t + u_t, \tag{4}$$

$$X_t = \sum_{j=1}^{J} C_j X_{t-j} + \varepsilon_t, \tag{5}$$

$$z_t = \mu^z + z_{t-1} + w_t \tag{6}$$

[7] A telltale sign of this in macroeconomics is that data and model outputs are often filtered before comparisons.

[8] We do not mean to imply that this model might not be misspecified along some dimensions. We think of it as a good description of many features of aggregate time series (more so than the economic theories we consider). One could enrich it to include a third vector of unobserved components that captures seasonal components, for example. We choose to use seasonally adjusted data in our empirical application instead.

where $u_t \overset{iid}{\sim} N(0, \Sigma_u)$ is a vector of measurement errors with a diagonal covariance matrix Σ_u[9], $\varepsilon_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon)$ and $w_t \sim N(0, \Sigma_w)$.

We assume that u_t, ε_t, and w_t are mutually independent. X_t can be interpreted as the cyclical component of the time series. The behavior of X_t is informed by economic theories via our construction of a mixture prior for C_j and Σ_ε. Notice that in the case where some model-based variables in the vector of latent states X_t have multiple measurements, the A matrix linking X_t to the observable indicators Y_t will have zeros and ones accordingly. z_t can be interpreted as the trend component of the time series. We allow for at most one random walk component per element of X_t so that various measurements of the same variables share the same low frequency behavior, as encoded in the selection matrix B.[10]

More general laws of motion for z_t can be incorporated, but in our specific application we use a random walk to capture low-frequency drift in inflation and the nominal interest rate.[11] We allow for a non-zero mean μ^z in the random walk equation to model variables that clearly drift such as log per capita real GDP in our application. If theories do have meaningful implications for trends of observables (as in our permanent income example in Section 5), our approach can easily be modified by dropping z_t from the model and directly using implications from the theories to allow for unit roots in the priors for equation (5) along the lines discussed below. In that case X_t would capture both the cycle and trend of our observables. μ captures differences in mean across various measurements of the same economic concept. Allowing for different measurements frees us from making somewhat arbitrary choices such as whether to base our analysis on CPI or PCE-based inflation only.

[9] We can allow this measurement error to be autocorrelated. This adds as many parameters to our system as there are observables because it adds one AR coefficient per observable as long as we model each measurement error as an autoregressive process of order 1. We do this as a robustness check in Section 5.

[10] Technically, we use a dispersed initial condition for z_0 and set the intercept in the measurement equation for one measurement per variable with a random walk trend to 0. That variable's intercept is then captured by z_0.

[11] A similar random walk assumption for inflation is commonly made in DSGE models, which in these models then imparts the same low frequency behavior in the nominal interest rate. The equilibrium models we consider in our New Keynesian empirical application do not have that feature.

Estimation via Gibbs Sampling

We can approximate the posterior via Gibbs sampling in three blocks with the mixture prior for the VAR coefficients $\gamma = (\beta, \Sigma_\varepsilon)$ in hand (the construction of which we describe below), where $\beta = vec(\{C_j\}_{j=1}^J)$. We focus here on an overview of the algorithm, details can be found in Appendix B. A Gibbs sampler draws parameters for a given block conditional on the parameters of all other blocks. One sweep through all blocks then delivers one new set of draws.

1. First, we draw the unobserved states X^T and z^T, which we estimate jointly. This can be achieved via various samplers for linear and Gaussian state space systems (Durbin and Koopman, 2012).

2. The second block consists of the parameters for the measurement equation μ, A, and Σ_u.[12] We use a Gaussian prior for μ and the free coefficients in A (if any - A can be a fixed selection matrix as in our empirical application, just as B) and an Inverse-Gamma prior for each diagonal element of Σ_u, which allows us to draw from the conditional posterior for those variables in closed form.

3. Finally, the VAR coefficients γ are drawn according to the algorithm for drawing from the mixture posterior outlined before (note that the conditional marginal likelihood that is needed for this algorithm is available in closed form for all conjugate priors we consider here). We have three options for drawing from a natural conjugate prior for γ when the marginal likelihood (conditional on the parameters in the other blocks) is known:

    (i) The Normal-inverse Wishart prior (Koop and Korobilis, 2010).

    (ii) A variant of that prior where Σ_ε is calibrated (fixed) a priori as in the classical Minnesota prior (see again Koop and Korobilis, 2010).

    (iii) The prior recently proposed by Chan (2021) that breaks the cross-equation correlation structure imposed by the first prior.

[12] Throughout we assume that the unobserved state vector X_t has a mean of zero. In our simulations from the equilibrium models we will demean all simulated data.

We use (iii) as our benchmark as it allows for a more flexible prior for γ while at the same time putting enough structure on the prior densities to make prior elicitation (i.e. the mapping from our simulated data to the prior) reasonably straightforward. We use prior (ii) for robustness checks - we find prior (i) to be too restrictive for actual empirical use. Another advantage of the prior structure introduced in Chan (2021) is that it is explicitly set up to be able to deal with a large number of observables, which means that our approach can also be used with a large dimensional X_t if an application calls for this.

2.2. *Simulating From Equilibrium Models* - Economic Theories

We assume that all economic models admit the following recursive representation:

$$X_{i,t} = \mathbf{F}_i(X_{i,t-1}, \varepsilon_{i,t}, \theta_i),\ \forall i = 1,...,K \tag{7}$$

where $\mathbf{F}_i$ is the mapping describing the law of motion of the state variables and $\varepsilon_{t,i}$ are the structural shocks in model i at period t. We focus on simulating demeaned data to be consistent with our state space models where the law of motion for X_t has no intercept.[13]

We require that we can simulate from (an approximation to) this recursive representation. The specific form of the approximation is not important per se, but should be guided by the economic question. If nonlinearities are important, researchers can use a nonlinear solution algorithm. We discuss the interplay of a nonlinear solution algorithm and our linear time series model in more detail in Section 3.1 below. Note that while solving models nonlinearly can be time consuming, this step of the algorithm can generally be carried out in parallel.

It is important to understand the role of simulating data X_t from a DSGE model to generate VAR priors. The intuition is that a DSGE model, conditional on priors on the deep parameters of the model, imposes very specific restrictions on the interactions among its model-based variables and, as such, it implies very specific contemporaneous and dynamic covariances as well

[13] We find it useful to assume that the dimension of $\varepsilon_{i,t}$ is at least as large as the dimension of X_t to minimize the risk of stochastic singularity in the next steps. One can add "measurement error" to $X_{i,t}$ to achieve this, for example.

as volatilities in the simulated sample. Therefore, a researcher can capture this information by examining the relationships in the simulated sample (e.g., they can infer dynamic correlations between inflation and output or the sign of the response of hours worked to a TFP shock). For our algorithm, we need N draws from the DSGE prior $f_i(\theta_i)$. For each of these draws, we simulate a sample of length T of the theory-based variables $X_{i,t}$, and then estimate a VAR on this simulated dataset. These N estimates of VAR parameters for each theory are then used to generate a prior for the VAR model, as we describe next.

2.3. Generating VAR Priors from Simulated Data

Mapping Economic Theories Into Macroeconomic Reality

As a reminder, we have K models with parameter vector θ_i and associated prior $f_i(\theta_i)$ we want to derive VAR priors from.[14] For a given n-dimensional vector of observables Y_t, we need a prior for the VAR coefficients γ and the residual covariance matrix Σ_u. We use a simulation-based approach to set our priors. This not only generalizes to non-linear DSGE models but also allows us to easily take parameter uncertainty into account.

1. To start, we simulate R datasets of length $T^{burn-in} + T^{final} = T$. We then discard the initial $T^{burn-in}$ periods for each simulation to arrive at a sample size of T^{final}.

    - We pick the number of simulated data sets R to be at least 2000 in our applications. We generally recommend to increase the number of simulations until the corresponding prior does not change anymore. Since simulating the equilibrium models and computing the prior parameters can be done largely in parallel, this approach is not time consuming. Our choice for T^{final} is 25 percent of the sample size of the

[14] The prior could be degenerate for some elements of θ if the researcher was interested in calibrating some parameters. The prior could also be informed by a training sample along the lines outlined by Del Negro and Schorfheide (2008).

actual data.[15]

- The choice of T^{final} implicitly governs how tight the variance of each mixture component is. If desired, a researcher can easily add an ad-hoc scaling factor to increase the variances of each mixture component.
- Notice that in the case where length of the simulated datatsets $T^{final} \to \infty$, VAR parameter uncertainty conditional on a specific value of the DSGE parameters θ_i will vanish, but since we allow for uncertainty about those DSGE parameters there can still be uncertainty about VAR parameters even if the simulated sample size grows very large.

2. For each model/mixture component, we choose the prior mean for the coefficients in the VAR to be the average VAR estimate across all simulations for that specific model. For the free parameters in the prior variances for γ, we set the elements equal to the corresponding elements of the Monte Carlo variance across simulations. Similarly, we use the Monte Carlo mean and variances to select the inverse-Gamma priors for the variances of the one-step ahead forecast errors (details can be found in Appendix C).

3. We use a VAR(2) for X_t in our empirical application. We choose a relatively small number of lags for parsimony, but show as a robustness check that using a VAR(4) instead gives very similar results, a practice we generally recommend.

2.4. An Illustration

To get a sense of how our approach works in practice, we consider a simple example with two key features. First, we assume that we directly observe X_t, so that the data is already cleaned of measurement error and stochastic trends, and we can thus focus only on the part of the Gibbs sampler with the mixture prior. Second, we assume that we only estimate *one parameter* in the

[15] For larger systems with either a larger dimension of X_t or a larger number of lags in the dynamics for X_t, we recommend to adjust this fraction for the standard reason that tighter priors are needed / preferred in larger systems. For example, later we use a robustness check where we use a VAR(4) instead of a VAR(2), which is our benchmark choice. In that case we double T^{final} to be 50 percent of the actual sample size (which is in line with the value chosen by Del Negro and Schorfheide, 2004) since we doubled the number of lags.

prior mixture block of the Gibbs sampler, which allows us to plot priors and posteriors easily. In the left panel of Figure 1, we consider an example where we have two priors (in blue), which we use to form the mixture prior (for simplicity we assume equal weights). These priors generally come from equilibrium models in our approach. What determines how the prior model weights are updated is the overlap between the likelihood and each mixture component, as can be seen in equation (3). While component 2 is favored in this example, component 1 still has substantial overlap with the likelihood and hence a non-negligible posterior weight. Note that even if the likelihood completely overlapped with component 2, component 1 could in general still receive non-zero weight because of the overlap between the two components of the prior mixture.

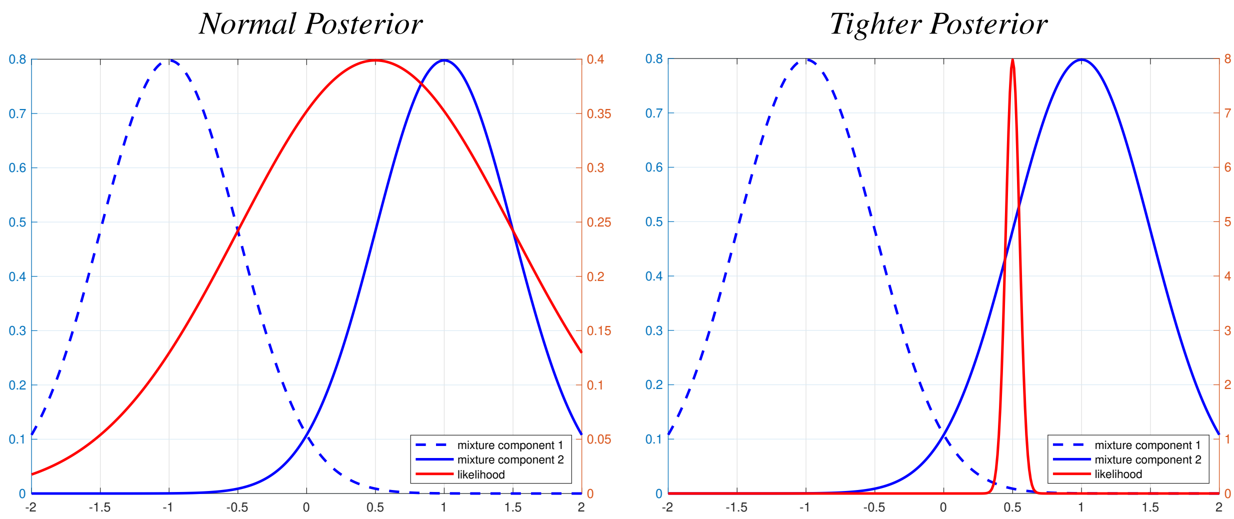

Figure 1: Components of Mixture Prior and Posterior.

We next turn to a scenario where the likelihood has less variance, as depicted in the right panel of Figure 1. What becomes apparent is that even as the posterior variance goes to zero (as it generally will in our applications with an increasing sample size) the model weights might still not become degenerate. This is not a flaw of our approach, but requires some discussion as to how we think about asymptotic behavior in this framework. Traditionally, in a Bayesian context one might think about asymptotic behavior as letting the sample size grow to infinity without changing the prior. In order to be asymptotically able to discriminate *with certainty* between theories with our approach, we should increase the sample size used to simulate data from the equilibrium models to form the prior. This will lead to the mixture components having less overlap and hence making discriminating between models easier. Note that this does *not* mean that the variance of the mixture prior will go to zero. Our benchmark approach automatically sets the size of the simulations that determine the VAR prior to be a constant fraction (25 percent) of the actual sample size.

Insofar as cleaning the data of measurement error and stochastic trends (as we do with our Gibbs sampler) removes outliers or makes them less severe, our approach is more robust to outliers in the data than using the observed data to compute the marginal likelihoods of the economic models directly. This is due to the fact that our model weights are based on marginal likelihood comparisons for the VAR of the unobserved cyclical component vector X_t, and thus the likelihood function for X_t does not move as much in response to outliers (i.e., when outliers are added to the data set) as the likelihood function of the theoretical model when such a theoretical model does not take into account all the measurement and trend issues that we do.[16]

3. Extensions

In this section we are going to consider two important extensions of our procedure. The first extension considers the case where a researcher is interested in using our procedure to incorporate information from nonlinear theories/models. The second shows how our methodology can be easily extended to the case where, instead of using fully specified economic theories, a researcher might be interested in imposing only certain implications of a theory while being agnostic about others. In Appendix D we also present a third extension, which describes how to estimate impulse responses via a data-driven approach to selecting among/averaging over identification schemes coming from different models/theories.

[16] By "likelihood function for X_t" we mean the density of X_t conditional on parameters, as depicted, for example, in Figure 1.

3.1. *Nonlinearities and the Choice of Variables*

While our time series model is linear, if the equilibrium models we study are solved using nonlinear solution methods and nonlinearities are possibly important for discriminating between theories, then our approach can exploit these nonlinearities. To highlight this point, consider a simplified version of our setup where, instead of a VAR, we use a univariate linear regression to discriminate among models:

$$x_{1,t} = \Xi x_{2,t} + \varepsilon_t$$

where $x_{1,t}$ and $x_{2,t}$ are demeaned variables simulated from an equilibrium model. We know that asymptotically $\hat{\Xi} = \frac{cov(x_{1,t}, x_{2,t})}{var(x_{2,t})}$, where *cov* and *var* are population moments. Under well known conditions, the regression coefficient on long simulations from the model will approach $\hat{\Xi}$. However, these population moments themselves will generally depend on the order of approximation used to solve and simulate the equilibrium model. It is *not* true that a first-order approximation and a non-linear solution method will generically deliver similar values for $\hat{\Xi}$, even though it is a regression coefficient in a linear regression and the decisions rules from a first-order approximation give the best linear approximation to the true nonlinear decision rules. If heterogeneity or movements in higher-order moments (such as standard deviations) are important and a feature of all equilibrium models that are studied, then measures of cross-sectional dispersion or higher-order moments can be included in the time series model if data on these moments are available. We can then think of the time series model as a linear approximation to the joint dynamics of aggregate variables and these higher-order or cross-sectional moments.[17]

3.2. *Incorporating Only Some Aspects of Theories*

Our approach allows researchers to include only some aspects of economic theories while being agnostic about others by using structural VARs instead of fully specified economic theories to form priors. This extension builds on Baumeister and Hamilton (2019), who show how to map beliefs about certain aspects of economic theories into a prior for a structural VAR. This

[17] Measures of higher order moments are, for example, commonly introduced in linear time series models to study the effects of uncertainty shocks - see Bloom (2009).

prior information may come from simple or more articulated theoretical models and can be summarized as a set of restrictions on the structural VAR matrices. For instance, prior beliefs about the magnitude or sign of the impulse responses of a dependent variable to a shock ("oil production does not respond to the oil price, on impact" or "a TFP shock leads to increase in hours worked") can be cast as priors on the impact matrix of the structural VAR model. For other features of the data, one can use non-informative or standard priors that are not directly linked to a fully specified economic theory. Aruoba et al. (2021) follow a similar logic in their structural VAR model in which an occasionally-binding constraint generates censoring in one of the dependent variables. In their application the censored variable is the nominal interest rate constrained by the effective lower bound and prior beliefs about directions of impulse responses are based on a simple New Keynesian DSGE model.

Our framework can be used to embrace several of such "theories" by using a modified version of our algorithm. To see this, first consider the following structural VAR model, which represents one of our "theories":

$$\mathscr{A}_{i,0} X_{i,t} = \mathscr{A}_{i,1} X_{i,t-1} + \nu_{i,t}, \tag{8}$$

where $X_{i,t}$ is an $(n_i \times 1)$ vector of variables in model i, $\mathscr{A}_{i,0}$ an $(n_i \times n_i)$ matrix describing contemporaneous structural relations, $X_{i,t-1}$ a $(k_i \times 1)$ vector encompassing a constant and m_i lags of $X_{i,t}$ (and, thus, $k_i = m_i n_i + 1$) and $\nu_{i,t} \sim \mathscr{N}(0, \Omega_{i,\nu})$ is an $(n_i \times 1)$ vector. As in Baumeister and Hamilton (2019), prior information about $\mathscr{A}_{i,0}$ in model i would be represented by the prior density $f_i(\mathscr{A}_{i,0})$ and may speak to individual elements of $\mathscr{A}_{i,0}$ (e.g., restrictions on contemporaneous relationships) or to its nonlinear combinations, such as elements of $\mathscr{A}_{i,0}^{-1}$ (e.g., restrictions on impulse responses). Priors on the other parameters could be informed by either economic theories or standard priors in the VAR literature (such as the Minnesota prior for $\mathscr{A}_{i,1}$ used in Baumeister and Hamilton, 2019).

To apply our procedure to such a framework, one simply needs to draw the structural VAR parameters from their prior, simulate data $X_{i,t}$ from model (8) conditional on these parameters, and then proceed with generating priors for the VAR parameters γ, before finally moving on to the "Macroeconomic Reality" step in our procedure (described in Section 2 and illustrated in Figure A-1 of Appendix A).

4. Some Monte Carlo Examples

To get a sense of how our approach performs and how it relates to standard measures of fit, we present a series of examples and associated Monte Carlo simulations. A first set of simulations highlights that our approach automatically provides model averaging or model discrimination based on the data-generating process. We do this using a setup with simple statistical models. Afterwards, we study an economic example where the models available to the researcher are misspecified in economically meaningful ways. Another Monte Carlo example, which highlights the trade-off between small and large models and the role played by the choice of observables in that trade-off, can be found in Appendix F.

4.1. Model Averaging and the Role of Measurement Error

To show that our approach can be used not only to discriminate between models but also to optimally combine them when they are all useful representations of the data, we now consider a simulation exercise. Here we simulate one sample per specification for the sake of brevity, but the random seed is fixed across specifications, so the innovations in the simulated data are the same across simulates samples (only the endogenous propagation changes). In this specification, we study 200 observations of a scalar time series y_t. We consider two models:

Model 1. *Less Persistence*

$$y_t = 0.7 y_{t-1} + e_t, \tag{9}$$

where $e_t \overset{iid}{\sim} N(0,1)$, and

Model 2. *More Persistence*

$$y_t = 0.9 y_{t-1} + e_t, \tag{10}$$

where again $e_t \overset{iid}{\sim} N(0,1)$. We consider three data-generating processes: One where the less persistent model is correct, one where the more persistent model is correct, and one where the

DGP switches from the first model to the second in period 101 (the middle of the sample).[18] Table 1 shows that if models fit (part of) the data well, our approach will acknowledge this, as both models receive basically equal weight in the case of the third DGP, whereas the correct model dominates in the first two DGPs.

Table 1: Posterior Mean of Model Weights, Second Simulation Exercise.

Data-Generating Process	(Average Posterior) Weight of Model 1	Weight of Model 2
Model 1 is correct	0.89	0.11
Model 2 is correct	0.17	0.83
Switch in $t = T/2$	0.48	0.52

In this setup we can also assess the role of measurement error further. While measurement error can make it harder to discriminate among models in theory, our measurement error is restricted to be i.i.d. When we redo our analysis, but now introduce extreme measurement error (where for simplicity we fix the measurement error variance to be 1), the results are basically unchanged. To give one example, if model 2 is the correct model and we allow for such measurement error in our estimated model (measurement error that is absent from the data-generating process), the model probability for model 2 is 0.91. Measurement error does not significantly move the estimated model weights because the difference across the two models is that model 2 is more persistent, which the iid measurement error cannot mask.

4.2. Misspecified Frictions in a DSGE Model

We now turn to the following scenario: Suppose the data-generating process is a LARGE equilibrium model with a substantial number of frictions. We simulate data from this model, but then compare two misspecified models: One model where one friction is turned off only, and another, smaller, model where other frictions are missing, but the one friction the first model gets wrong is actually present. As a laboratory, we use the Schmitt-Grohe and Uribe (2012)

[18] We use the standard Normal-inverse Wishart prior. We shut down the random walk component and assume we observe a measurement of y_t that is free of measurement error in the first three exercises.

real business cycle (RBC) model.[19] This is a real model with a rich set of frictions to match US data. Schmitt-Grohe and Uribe (2012) study news shocks, but for simplicity we turn off these shocks in our exercises (so they are not present in neither the DGP nor the models we want to attach weights to).

Our DGP adds one friction to the standard Schmitt-Grohe and Uribe (2012) model: A time varying discount factor, where time variation can be due to time variation in the capital stock or an exogenous shock. The specification we use for this time variation is due to Canova et al., 2020. Time variation in the discount factor has become a common tool to model, for example, sudden shifts in real interest rates - see for example Bianchi and Melosi (2017). The first model we consider is equal to the DGP except that it lacks time variation in the discount factor. The second model features this time variation, but lacks other frictions present in the DGP: Habits, investment adjustment costs, varying capacity utilization, and mark-up shocks. We simulate 200 Monte Carlo samples of length 250. The parameter values of each model are calibrated for each simulated data set (so the priors $f_i(\theta_i)$ are degenerate, which makes it easier for us to discuss misspecification).[20] As observables we use consumption, hours, and investment.

When we carry out this exercise, the result is not surprising: The larger model which only has one misspecification is clearly preferred by the data (average model weight of 1). What we are interested in, is how our model weights behave as we make the models closer to each other and if there are substantial differences relative to Bayesian model probabilities. To do so, we exploit that fact that each of the fractions that the smaller model is missing is governed by one parameter. We now reintroduce two of these frictions, habits in consumption and investment adjustment costs, but not using the true parameter values, but a common fraction x (with $x < 1$) of the true value. Table 2 shows the results for selected values of x. As we increase x, our approach realizes that both models provide useful features to match the data. Bayesian model probabilities of the equilibrium models themselves, on the other hand, suggest that only the

[19] In Appendix F, we use a New Keynesian model in another Monte Carlo exercise. In that exercise, we also extend our framework to allow for additional exogenous regressors in our VAR for X_t so that it can be used, for example, to focus on implications for a single variable such as the nominal interest rate.

[20] We discuss details of the calibration in Appendix E.

larger model is useful. Our approach is more cautious in that it tends to give positive weights to all available models, similarly to the opinion pools literature (Geweke and Amisano, 2011). This feature will also be present in our empirical examples later.[21]

Table 2: Model Weights and Model Probabilities.

Model Specification (value of x)	Our Approach	Bayesian Model Probability
0.25	0.21	0.00
0.33	0.46	0.00
0.40	0.61	0.00
0.55	0.78	0.00

5. Does Heterogeneity Matter for Aggregate Behavior?

A key question for anyone trying to write down a macroeconomic model is whether to include household heterogeneity. A traditional answer, using results from Krusell and Smith (1998), is that household heterogeneity might not matter for aggregate outcomes as much as we would think. In this section we highlight that this is not a general result by using two empirical applications: (i) a permanent income example, where we use aggregate US data on real income and consumption to distinguish between a representative agent permanent income model and a version of the same model that also has hand-to-mouth consumers, and (ii) a stylized three equation New Keynesian model where we again contrast the representative agent version with a version that also has hand-to-mouth households.

These two examples share similarities (such as the use of hand-to-mouth consumers to introduce a stylized notion of household heterogeneity), but they also differ in important aspects: The permanent income example uses theories that have implications for trends. We therefore shut down the trends in our unobserved components model and instead allow for unit roots in the VAR for X_t.[22] The New Keynesian application instead uses theories that only have implications

[21] In Appendix I, we provide a more detailed discussion of the differences between standard Bayesian model probabilities and our approach, in particular focusing on why our approach will likely lead to less extreme model weights.

[22] The first application also shuts down measurement error in the observation equation of our unobserved components model.

for cyclical behavior.

Permanent Income Models

We borrow the representative household version of our linear-quadratic permanent income model from Inoue et al. (2020). The two-agent version adds hand-to-mouth consumers that cannot invest in the riskless bond that is available to the other households. The parameters of each model are calibrated to fit our data on real per-capita income and consumption in the US.[23] We relegate details to Appendix G, but two features of the calibration are worth pointing out. First, we fix the fraction of hand-to mouth consumers at 0.25, a standard value in the literature. Second, the other parameters are calibrated to features of the income and consumption processes, but not the comovement of these two variables. Our approach thus exploits differences in the comovement between these variables implied by the two theories to distinguish between the models.

The top two lines of Table 3 show that household heterogeneity is preferred by the data. However, our approach does call for non-denegerate weights, whereas standard Bayesian model probabilities would put all weight on the two-agent model. We have already highlighted this cautious behavior by our approach in the Monte Carlo exercises. To convince readers that the representative agent model could be preferred by our algorithm if the data called for this, we also carry out a Monte Carlo exercise where the representative agent (RA) model is the data-generating process. We simulate 200 samples of length 100 and report the average posterior mean of our model weights as well as the average Bayesian model probability from this exercise in Table 3.

A New Keynesian Model of Household Heterogeneity

In recent work, Debortoli and Galí (2018) explore the implications of household heterogeneity for aggregate fluctuations. To do so, they depart from the Representative Agents New Keynesian (RANK) model and build a Two Agent New Keynesian (TANK) model featuring two

[23] Details can be found in Appendix G.

Table 3: Results, Permanent Income Models.

Model and Data Used	Our Approach	Bayesian Model Probabilities
One household, US data	0.35	0.00
Two households, US data	0.65	1.00
One household, RA model is DGP	0.75	0.69
Two households, RA model is DGP	0.25	0.31

types of households, namely "unconstrained" and "constrained" households, where the type is respectively determined by whether a household's consumption satisfies the consumption Euler equation. A constant share of households is assumed to be constrained and to behave in a "hand-to-mouth" fashion in that they consume their current income at all times.

Their framework shares a key feature with Heterogeneous Agents New Keynesian models (HANK): At any given point in time there is a fraction of households that faces a binding borrowing constraint and thus cannot adjust their consumption in response to changes in interest rates or any variable other than their current income. Relative to HANK models, the TANK framework offers greater tractability and transparency , but it comes at the cost of assuming a more stylized form of household heterogeneity. Nonetheless, Debortoli and Galí (2018) also show that TANK models approximate the aggregate output dynamics of a canonical HANK model in response to monetary and non-monetary shocks reasonably well.

We use versions of these models to simulate data on log output, hours worked, real interest rate and productivity.[24] The model equations as well as the priors over the deep parameters of the DSGE model can be found in Appendix H.1 and Appendix H.2. All DSGE priors (the $f_i(\theta_i)$ distributions) are common across the two models except the prior for λ, the fraction of constrained households in the two-agent model. While this fraction is set to 0 in the RANK model, we use a truncated normal distribution which is truncated to be between 0.1 and 0.3 (the underlying non-truncated distribution has mean 0.2 and standard deviation 0.1). The priors for

[24] The only differences relative to Debortoli and Galí (2018) are that we use introduce a cost-push shock instead of their preference shifter (to be comparable to most other small scale New Keynesian models), and that we use a backward-looking monetary policy rule, which we found made the OLS-based VAR estimates based on simulated data more stable.

those parameters that are not informed by the DSGE model are described in detail in Appendix H.3. The prior model weights are 0.5 each. As observables, we use quarterly log of per capita real GDP and GDI as measures of output, annualized quarterly CPI and PCE inflation, and the Federal Funds rate and the three months T-bill rate for the nominal interest rate. Our sample starts in 1970 and ends in 2019. More details on the data can be found in Appendix H.4.

Both of these theories are very much stylized: They disregard trends in nominal and real variables, the one theory that allows for heterogeneity does so in a stylized fashion, and the models will be approximated using log-linearization, thus disregarding any non-linearities. Nonetheless, we will see below that the TANK model provides a better fit to the data series we study.

We introduce random walk components for all three variables (output, inflation, and nominal interest rates). The different measurements for the same variables are restricted to share the same low-frequency random-walk components and the same cyclical components, but they can have different means and different high-frequency components. We use a larger prior mean for the innovation to the random walk in output compared to inflation and nominal interest rates to account for the clear trend. We allow for independent random walk components in inflation and the nominal interest rate to be flexible, but one could restrict those variables to share the same trend (as would be implied by equilibrium models with trend inflation described by, for example, Cogley and Sbordone, 2008). In Table 4, we show the resulting model probabilities.

Table 4: Posterior Mean of Model Weights, RANK vs. TANK. Benchmark Specification and Various Robustness Checks.

	RANK	**TANK**
Chan's prior	0.06	0.94
Chan's prior with Wu/Xia shadow rate	0.12	0.88
N-IW prior	0.00	1.00
T^{final} equal to half the actual sample size	0.01	0.99
VAR(4)	0.13	0.87
Autocorrelated Measurement Error	0.08	0.92

They show that in the current specification, the TANK model has a clear advantage over its representative agent counterpart in fitting standard aggregate data often used to estimate New

Keynesian models. To assess robustness, we carry out two robustness checks: (i) we replace the two measures of the nominal interest rate with the Wu and Xia (2016) shadow interest rate (we do this to reduce model misspecification since the model lacks the nonlinear features to deal with the zero lower bound period), and (ii) we use a common Normal-inverse Wishart Prior instead of the Chan (2021) prior. As highlighted in Table 4, our main finding of TANK superiority is robust. In fact, we find that with a Normal-inverse Wishart prior TANK is even more preferred. We think this is most likely an artifact of the strong restrictions implied by that prior, leading us to prefer Chan's prior instead.

To assess robustness of our results with respect to some of the specification choices we have made, we now vary the lag length in the VAR for the cyclical component X_t, modify T^{final}, the length of the simulated data series that are used to inform our prior, and allow for autocorrelated measurement error. For the first robustness check, we set this sample size to half the actual sample size. Not surprisingly, the results are even stronger than in our benchmark case. We also check robustness with respect to the number of lags included in the VAR for X_t. If we use a VAR(4) instead of our VAR(2) benchmark, we still see that the TANK model is strongly preferred by the data. Finally, if we allow for autocorrelated measurement error, the TANK model is still substantially preferred by the data.

6. Conclusion

We propose an unobserved components model that uses economic theories to inform the prior of the cyclical components. If theories are also informative about trends or even seasonal fluctuations, our approach can be extended in a straightforward fashion. Researchers can use this framework to update beliefs about the validity of theories, while at the same time acknowledging that these theories are misspecified. Our approach inherits benefits from standard VAR and unobserved components frameworks, while at the same time enriching these standard approaches by allowing researchers to use various economic theories to inform the priors and to learn about the fit of each of these theories.

References

Amisano, G. and J. Geweke (2017). Prediction Using Several Macroeconomic Models. *The Review of Economics and Statistics 99*(5), 912–925.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016). Improving GDP Measurement: A Measurement-Error Perspective. *Journal of Econometrics 191*(2), 384–397.

Aruoba, S. B., M. Mlikota, F. Schorfheide, and S. Villalvazo (2021). SVARs with Occasionally-Binding Constraints. Working Paper 28571, National Bureau of Economic Research.

Baumeister, C. and J. D. Hamilton (2019). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *American Economic Review 109*(5), 1873–1910.

Bianchi, F. and L. Melosi (2017). Escaping the Great Recession. *American Economic Review 107*(4), 1030–1058.

Bilbiie, F. O. (2018). Monetary Policy and Heterogeneity: An Analytical Framework. CEPR Discussion Papers 12601.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica 77*(3), 623–685.

Box, G. and N. Draper (1987). *Empirical Model-Building and Response Surfaces*. Wiley Series in Probability and Statistics. New York: Wiley.

Canova, F. (1994). Statistical Inference in Calibrated Models. *Journal of Applied Econometrics 9*(S1), S123–S144.

Canova, F. (2014). Bridging DSGE Models and the Raw Data. *Journal of Monetary Economics 67*, 1–15.

Canova, F., F. Ferroni, and C. Matthes (2020). Detecting and Analyzing the Effects of Time-Varying Parameters in DSGE Models. *International Economic Review 61*(1), 105–125.

Canova, F. and C. Matthes (2021). A Composite Likelihood Approach for Dynamic Structural Models. *The Economic Journal 131*(638), 2447–2477.

Chan, J. C. C. (2021). Asymmetric Conjugate Priors for Large Bayesian VARs. *Quantitative Economics*. forthcoming.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007). Business Cycle Accounting. *Econometrica 75*(3), 781–836.

Chib, S. and B. Ergashev (2009). Analysis of Multifactor Affine Yield Curve Models. *Journal of the American Statistical Association 104*(488), 1324–1337.

Chib, S. and R. C. Tiwari (1991). Robust Bayes Analysis in Normal Linear Regression with an Improper Mixture Prior. *Communications in Statistics - Theory and Methods 20*(3), 807–829.

Cogley, T. and A. M. Sbordone (2008). Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve. *American Economic Review 98*(5), 2101–2126.

Cogley, T. and R. Startz (2019). Robust Estimation of ARMA Models with Near Root Cancellation. In *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A*, Volume 40A of *Advances in Econometrics*, pp. 133–155. Bingley: Emerald Publishing Ltd.

Debortoli, D. and J. Galí (2018). Monetary Policy with Heterogeneous Agents: Insights from TANK Models. mimeo.

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (1996). A Bayesian Approach to Calibration. *Journal of Business & Economic Statistics 14*(1), 1–9.

Del Negro, M., R. B. Hasegawa, and F. Schorfheide (2016). Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance. *Journal of Econometrics 192*(2), 391–405.

Del Negro, M. and F. Schorfheide (2004). Priors from General Equilibrium Models for VARs. *International Economic Review 45*(2), 643–673.

Del Negro, M. and F. Schorfheide (2008). Forming Priors for DSGE Models (and how it Affects the Assessment of Nominal Rigidities). *Journal of Monetary Economics 55*(7), 1191–1208.

Den Haan, W. J. and T. Drechsel (2021). Agnostic Structural Disturbances (ASDs): Detecting and Reducing Misspecification in Empirical Macroeconomic Models. *Journal of Monetary Economics 117*, 258–277.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Fedoroff, A. (2016). *A Geometric Approach to Eigenanalysis-with Applications to Structural Stability*. Doctoral thesis, School of Engineering Aalto University.

Fessler, P. and M. Kasy (2019). How to Use Economic Theory to Improve Estimators: Shrinking Toward Theoretical Restrictions. *The Review of Economics and Statistics 101*(4), 681–698.

Filippeli, T., R. Harrison, and K. Theodoridis (2020). DSGE-Based Priors for BVARs and Quasi-Bayesian DSGE Estimation. *Econometrics and Statistics 16*, 1–27.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis*. Boca Raton: CRC Press.

Geweke, J. (2010). *Complete and Incomplete Econometric Models*. Princeton: Princeton University Press.

Geweke, J. and G. Amisano (2011). Optimal Prediction Pools. *Journal of Econometrics 164*(1), 130–141.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics 97*(2), 436–451.

Giannone, D., M. Lenza, and G. E. Primiceri (2019). Priors for the Long Run. *Journal of the American Statistical Association 114*(526), 565–580.

Granger, C. (1999). *Empirical Modeling in Economics: Specification and Evaluation*. Cambridge: Cambridge University Press.

Gregory, A. W. and G. W. Smith (1991). Calibration as Testing: Inference in Simulated Macroeconomic Models. *Journal of Business & Economic Statistics 9*(3), 297–303.

Ingram, B. F. and C. H. Whiteman (1994). Supplanting the 'Minnesota' Prior: Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors. *Journal of Monetary Economics 34*(3), 497–510.

Inoue, A., C.-H. Kuo, and B. Rossi (2020). Identifying the Sources of Model Misspecification. *Journal of Monetary Economics 110*, 1–18.

Ireland, P. N. (2004). A Method for Taking Models to the Data. *Journal of Economic Dynamics and Control 28*(6), 1205–1226.

Kaplan, G., B. Moll, and G. L. Violante (2018). Monetary Policy According to HANK. *American Economic Review 108*(3), 697–743.

Koop, G. and D. Korobilis (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends®in Econometrics 3*(4), 267–358.

Krusell, P. and A. A. Smith (1998). Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Lee, P. M. (2012). *Bayesian Statistics - An Introduction*. Chichester, Hoboken: John Wiley & Sons Ltd.

Schmitt-Grohe, S. and M. Uribe (2012). What's News in Business Cycles. *Econometrica 80*(6), 2733–2764.

Schorfheide, F. (2000). Loss Function-Based Evaluation of DSGE models. *Journal of Applied Econometrics 15*(6), 645–670.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica 48*(1), 1–48.

Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review 97*(3), 586–606.

Smith, A A, J. (1993). Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions. *Journal of Applied Econometrics 8*(S1), S63–S84.

Villani, M. (2009). Steady-State Priors for Vector Autoregressions. *Journal of Applied Econometrics 24*(4), 630–650.

Waggoner, D. F. and T. Zha (2012). Confronting Model Misspecification in Macroeconomics. *Journal of Econometrics 171*(2), 167–184.

Watson, M. W. (1993). Measures of Fit for Calibrated Models. *Journal of Political Economy 101*(6), 1011–1041.

Wu, J. C. and F. D. Xia (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking 48*(2-3), 253–291.

Appendix A. A Bird's Eye View of Our Approach

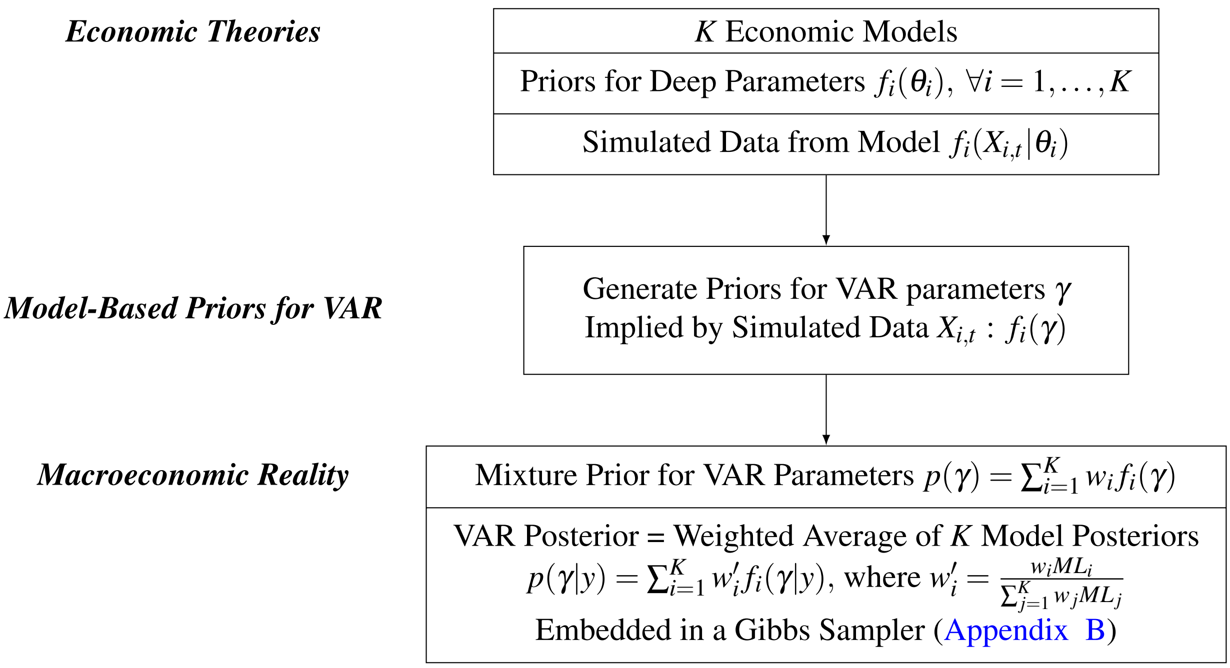

Figure A-1: From Economic Theories to Macroeconomic Reality via Model-Based Priors.

Appendix B. The Gibbs Sampler

The Gibbs sampler draws from the following conditional posterior distributions. First, define $\tilde{X}_t = (X_t', z_t')'$ and rewrite the model as

$$\begin{aligned} Y_t &= \mu + \tilde{A}\tilde{X}_t + u_t \\ \tilde{X}_t &= \mu_{\tilde{X}} + \tilde{C}X_{t-1} + \tilde{\varepsilon}_t. \end{aligned}$$

where $\tilde{A} = (A, B)$, $\tilde{C} = \begin{pmatrix} C & 0 \\ 0 & I_m \end{pmatrix}$, $\mu_{\tilde{X}} = (0, \mu_z')'$, $\tilde{\varepsilon}_t = (\varepsilon_t', w_t')'$.[25]

[25] We assume here w.l.o.g. that X_t follows a VAR(1) such that C is the only coefficient matrix. Otherwise, C is simply the coefficient matrix in companion form. $\tilde{X}_t$ and $\tilde{\varepsilon}_t$ have to be redefined accordingly.

1. Draw $\tilde{X}_t$ conditioned on $\mu, \mu_{\tilde{X}}, A, B, C, \Sigma_u, \Sigma_\varepsilon, \Sigma_w$ using the Carter and Kohn (1994) algorithm.

2. Σ_w is diagonal. Draw each diagonal element $\sigma_{w,i}^2$ conditioned on z_i^T from the inverse Gamma distribution $IG(\alpha_i^w, \beta_i^w)$ for $i = 1, \ldots, M$ with $\alpha_i^w = \alpha_i^{w,0} + \frac{T}{2}$ and $\beta_i^w = \beta_i^{w,0} + \frac{\Sigma\left(z_{i,t}-\mu_{z,i}-z_{i,t-1}\right)^2}{2}$, where $\alpha_i^{w,0}$ and $\beta_i^{w,0}$ are prior hyperparameters of $IG\left(\alpha_i^{w,0}, \beta_i^{w,0}\right)$.

3. Draw $\mu_{z,i}$ from the Normal distribution $N\left(\mu_{z,i}^*, V_{z,i}^*\right)$ with $V_{z,i}^* = \frac{1}{\frac{1}{V_{z,i}^0}+\frac{T}{\sigma_{z,i}^2}}$ and $\mu_{z,i}^* = V_{z,i}^*\left(\frac{\mu_{z,i}^0}{V_{z,i}^0} + \frac{\Sigma\left(z_{i,t}-z_{i,t-1}\right)}{\sigma_{z,i}^2}\right)$, where $\mu_{z,i}^0$ and $V_{z,i}^0$ are prior hyperparameters of $N\left(\mu_{z,i}^0, V_{z,i}^0\right)$.

4. Define $\tilde{Y}^T = Y^T - \tilde{A}\tilde{X}_t$. Σ_u is diagonal. Draw each diagonal element $\sigma_{u,j}^2$ from the inverse Gamma $IG\left(\alpha_j^u, \beta_j^u\right)$ for $j = 1, .., N$ with $\alpha_j^u = \alpha_j^{u,0} + \frac{T}{2}$ and $\beta_j^u = \beta_j^{u,0} + \frac{\Sigma\left(\tilde{y}_{j,t}-\mu_j\right)^2}{2}$, where $\alpha_j^{u,0}$ and $\beta_j^{u,0}$ are prior hyperparameters of $IG\left(\alpha_j^{u,0}, \beta_j^{u,0}\right)$.

5. Draw μ_j from the Normal distribution $N\left(\mu_j^*, V_j^*\right)$ with $V_j^* = \frac{1}{\frac{1}{V_j^0}+\frac{T}{\sigma_{u,j}^2}}$ and $\mu_j^* = V_j^*\left(\frac{\mu_j^0}{V_j^0} + \frac{\Sigma\tilde{y}_{j,t}}{\sigma_{u,j}^2}\right)$, where μ_j^0 and V_j^0 are prior hyperparameters of $N\left(\mu_j^0, V_j^0\right)$.

6. Compute posterior weights w_k' given draws of X^T for $k = 1, .., K$ based on the analytical marginal likelihood (see either Giannone et al., 2015 or Chan, 2021). Draw a model indicator δ based on w_k'.

7. For $\delta = k$, draw C and Σ_ε from the conjugate posterior associated with prior k.

8. Repeat 1-7 L times.

Appendix C. Mapping Simulated Data into Priors

For each model k, we simulate R data sets of length T^{final}. The OLS estimates of the VAR based on the simulated data form the basis of the prior for our empirical model. Two different options of priors are discussed below.

1. One prior option is the standard Normal- inverse Wishart natural conjugate prior. Let us recall the notation of the VAR conjugate prior

$$\begin{aligned} \Sigma &\sim IW(S, df) \\ \beta|\Sigma &\sim N(\hat{\beta}, \Sigma \otimes V) \end{aligned}$$

   where Σ is $M \times M$ matrix of residual covariance, β is a $\left(M + M^2 p\right) \times 1$ vector of VAR coefficients. Notice that the dimension of V is $(1 + Mp) \times (1 + Mp)$ which essentially defines the prior covariance of one equation of the VAR. The overall prior covariance is scaled by Σ.

   We set $\hat{\beta}$ equal to the average over OLS coefficient estimates of simulated data and df equal to the sample size of simulated data T^{final}. Let V_n denote the covariance over OLS coefficient estimates of simulated data. Furthermore, let Σ_n be the the average over OLS residual covariance estimates of simulated data. We set the prior location $S = \Sigma_n(df - n - 1)$. The main problem is that given Σ, V_n has more entries than unknowns in $\Sigma \otimes V$, hence the system is over-determined. We thus use a least square procedure to calibrate V. Following Fedoroff (2016), we can reformulate the problem as a linear system

$$vec\left(\Sigma_n \otimes V\right) = Avec(V) = vec(V_n),$$

   where Σ is replaced by Σ_n. We then solve for

$$vec(V) = \left(A'A\right)^{-1} A' vec(V_n).$$

A is a function of Σ taken from Lemma B.4 of Fedoroff (2016) (p.181).

2. Our benchmark approach uses the asymmetric conjugate prior by Chan (2021). The essential assumption is that VAR coefficients are independent across equations. Using the notation from Chan (2021), the Normal-inverse-Gamma prior for each equation i can be written as

$$\theta_i|\sigma_i^2 \sim N(m_i, \sigma_i^2 V_i), \qquad \sigma_i^2 \sim IG(\nu_i, S_i),$$

where $\theta_i = (\alpha_i', \beta_i')'$ is the collection of reduced form parameters β_i and elements of the impact matrix α_i. Prior means of β_i are set equal to the average over OLS coefficient estimates of simulated data. Prior means for α_i are set equal to zero. ν_i and S_i are calibrated to the mean and variance over OLS residual covariance estimates of simulated data of the associated equation. Let $\hat{\sigma}_i^2$ denotes the average of OLS residual variance estimates of equation i. V_i is assumed to be diagonal, where the variances of β_i are set to variance over OLS coefficient estimates of simulated data (scaled by $\hat{\sigma}_i^2$). Covariances of α_i are diagonal and set to $1/\hat{\sigma}_i^2$.

Appendix D. Data-Driven Averaging Over Identification Restrictions

Next we consider a scenario where for each model i we have an impulse response function $\mathscr{M}_i(\gamma)$, where γ is the vector of VAR coefficients. This theory-specific function (it is indexed by i) returns a matrix of impulse responses R for all variables X_t in the VAR and for horizons $h = 0, 1, ..., H-1$, where 0 is the impact horizon. We will need to put some structure on the $\mathscr{M}_i$ functions, namely that the associated identification restrictions either lead to set-identification or exact identification.[26] In other words, any identification restrictions imposed here do not influence the fit of the VAR models. Other than that there is still substantial freedom to choose the exact form of the $\mathscr{M}_i$ function for each model (for example, the choice of what sign restrictions

[26] We assume that if impulse responses for a given model are only set identified, $\mathscr{M}_i(\gamma)$ randomly selects one valid impulse response vector.

are imposed and for what horizons).

Remember the structure of our Gibbs sampler: We draw a model indicator i according to the implied model probabilities based on the marginal likelihoods of the VARs, and then conditional on that indicator i we draw VAR parameters. We can then add an additional step right after that generates a draw from $\mathscr{M}_i(\gamma)$ (we could also do this ex-post after the reduced-form estimation). We collect the resulting draws of the IRFs to approximate the posterior of the impulse responses. What does this procedure give us? It gives us a *data-driven approach to select among /average over identification schemes coming from different models*. The posteriors of the IRFs take into account uncertainty about identification schemes. If one theory gets almost all the probability weight, we will basically only use identification restrictions from that theory. We can also generalize the approach and allow $n_i = 1, \ldots, N_i$ identification schemes *per theory*, each associated with a prior probability $f_i^{n_i}$. After drawing a model indicator, we could then draw an identification scheme indicator according to those probabilities. While our approach would not allow use to update the probabilities for a given model, the resulting posteriors would take into account the uncertainty across these identification schemes for each model.[27]

Appendix E. Schmitt-Grohe and Uribe (2012) Monte Carlo Exercise: Calibration

The DGP is a modified version of Schmitt-Grohe and Uribe (2012) model with the time varying discount factor. All model features containing news shocks are removed. All parameters except parameters governing the time varying discount factor are fixed to MLE estimates of Schmitt-Grohe and Uribe (2012).

[27] Since the models in our empirical application provide similar impulse responses to the shocks in the models we do not compute impulse responses there. Nonetheless, we include this section because we think it is useful for readers to be aware of this additional application of our approach.

The time variation in the discount factor β_t is modelled following Canova et al. (2020):

$$\Theta_t = \beta_t / \beta_{t-1} \tag{E-1}$$

$$\Theta_{t+1} = \left[\Theta_u - (\Theta_u - \Theta_l)\, e^{-\phi_a (K_t - K)}\right] + \left[\Theta_u - (\Theta_u - \Theta_l)\, e^{\phi_b (K_t - K)}\right] + \sigma_u * U_{\theta,t+1} \tag{E-2}$$

All relevant structural parameters of both first (large) and second (small) model are fixed to DGP values. For the parameters of the time varying discount factor, we set $\phi_a = 0.1$, $\phi_b = 0.4$, $\Theta_u = 0.995$, $\Theta_l = 0.99$ and $\sigma_u = 1$. Persistence and standard deviations of all shocks are calibrated to match variances, first and fourth order autocorrelations of consumption, hours and investment of the DGP. In the Monte Carlo simulation, we set the lag length of the VAR to 2 and the measurement equations only includes intercepts, where the prior means are set to 0 and variances to 100. R, the number of simulated data sets, is set to 5000.

Appendix F. New Keynesian Monte Carlo Exercise

To highlight how our approach could be extended to focus on the implications on a specific variable, we now use another Monte Carlo example. Our DGP is the Smets and Wouters (2007) New Keynesian Model. The two models we consider are: (i) the Smets and Wouters (2007) model, but with a misspecified monetary policy rule that does not react to the output gap, and (ii) a three-equation New Keynesian model (Del Negro and Schorfheide, 2008) with the correctly specified policy rule. We re-estimated all three models on the same dataset (using output growth, inflation and the nominal rate) - a subset of the original Smets and Wouters (2007) dataset. Parameters for the DGP and the degenerate prior distributions are then set to the posterior means of those estimations.

As a benchmark, we first use our standard approach. We simulate 200 Monte Carlo samples of inflation, nominal interest rates, and the output gap of length 250 each. We drop measurement error and trends from our unobserved component model for the sake of simplicity. We set the number of simulated data sets $R = 5000$ and the total number of Gibbs sampler draws is 20000.

Similarly to what we found for the Schmitt-Grohe and Uribe (2012) exercise, the larger model is still preferred by the data, with an average model posterior mean of the model weight across samples of 0.99. We use this example to highlight how our approach could be extended if a researcher was interested in implications for one variable specifically (say, the nominal interest rate in this example). While one could use our approach directly in that scenario, it might not be very appealing because one would be left with estimating an autoregressive process for that variable only, disregarding any interactions with other variables. We thus extend our approach and replace the VAR for X_t with a VAR with additional exogenous variables:

$$X_t = \sum_{j=1}^{J} C_j X_{t-j} + \sum_{l=0}^{L} M_l Z_{t-l} + \varepsilon_t$$

where Z_t is a vector of exogenous observable variables.[28] In our application, we focus on the nominal interest rate, use $J = 1$ and $L = 0$ lags, and set $Z_t = [\pi_t \; y_t]'$, so that this equation has the same form as the true policy rule. Because the smaller model does have the correct form of the policy rule, this smaller model now gets an average posterior mean weight of 1.

Appendix G. Consumption and Liquidity Constraints

The representative household version of our linear-quadratic permanent income model used in Section 5 borrows from Inoue et al. (2020). The representative agent model is given by:

$$c_t = \frac{r}{r+1} a_t + \left(y_t^P + \frac{r}{1-\rho+r} y_t^T\right) \tag{G-1}$$

$$y_t^T = \rho y_{t-1}^T + e_{1t} \tag{G-2}$$

$$y_t^P = \gamma + y_{t-1}^P + e_{2t} \tag{G-3}$$

$$y_t = y_t^T + y_t^P \tag{G-4}$$

$$a_{t+1} = (1+r)(a_t + y_t - c_t) \tag{G-5}$$

[28] We assume for simplicity here that Z_t has mean 0.

The two-agent version adds hand-to-mouth consumers that cannot invest in the riskless bond that is available to the other households:[29]

$$c_t^1 = \frac{r}{r+1} a_t + (y_t^P + \frac{r}{1-\rho+r} y_t^{1,T}) \tag{G-6}$$

$$c_t^2 = y_t^2 \tag{G-7}$$

$$c_t = \omega c_t^1 + (1-\omega) c_t^2 \tag{G-8}$$

$$y_t^{1,T} = \rho^1 y_{t-1}^{1,T} + e_{1t} \tag{G-9}$$

$$y_t^{2,T} = \rho^2 y_{t-1}^{2,T} + e_{2t} \tag{G-10}$$

$$y_t^P = \gamma + y_{t-1}^P + e_{3t} \tag{G-11}$$

$$y_t^1 = y_t^{1,T} + y_t^P \tag{G-12}$$

$$y_t^2 = y_t^{2,T} + y_t^P \tag{G-13}$$

$$a_{t+1} = (1+r)(a_t + y_t^{1,T} + y_t^P - c_t^1) \tag{G-14}$$

$$y_t = \omega y_t^1 + (1-\omega) y_t^2 \tag{G-15}$$

All shocks in both models are denoted by e and are Gaussian and independent of each other. We use real disposable personal income per capita (A229RX0) and real personal consumption expenditures per capita (A794RX0Q049SBEA) provided by FRED as observables. The data range from 1985Q1 to 2005Q4. We set γ, the growth rate of income, equal to the average quarterly income growth of the data. The (quarterly) real interest rate is set to 0.005 and ω, the fraction of unconstrained household, to 0.75. The remaining parameters of both models are calibrated to match standard deviations of income and consumption growth of the data and are given in table G-1.

The time series model uses log income and log consumption in levels as observables Y_t. The VAR length is set to 2 and contains intercepts in each equation. The measurement equations do not contain intercepts, low frequency components or measurement errors. Specifically, the state-space model takes the following form:

[29] Superscript 2 denotes variables belonging to constrained agents.

$$
\begin{aligned}
Y_t &= X_t \\
X_t &= \mu + C_1 X_{t-1} + C_2 X_{t-2} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon)
\end{aligned}
$$

Mixture-priors are employed on μ, C_1, C_2,and Σ_ε, whose prior hyperparameters are informed by VAR(2) estimates of simulated data.

Table G-1: Calibration, Permanent Income Example

Parameter	Value	Target
r	0.0050	quarterly real interest rate
γ	0.0051	average quarterly real income growth
		Representative Agent Model
ρ	0.2409	standard deviations of income and consumption growth
$100 \times \sigma_{e_1}$	0.2995	standard deviations of income and consumption growth
$100 \times \sigma_{e_2}$	0.4636	standard deviations of income and consumption growth
		Heterogeneous Agent Model
ω	0.75	standard value in the literature
ρ^1	0.4998	standard deviations of income and consumption growth
ρ^2	0.4999	standard deviations of income and consumption growth
$100 \times \sigma_{e_1}$	0.4905	standard deviations of income and consumption growth
$100 \times \sigma_{e_2}$	0.0161	standard deviations of income and consumption growth
$100 \times \sigma_{e_3}$	0.4805	standard deviations of income and consumption growth

Appendix H. Debortoli and Galí (2018): RANK vs. TANK

The Debortoli and Galí (2018) model used in Section 5 is described by the following set of equations, where we provide the description of the endogenous and exogenous variables in tables H-2 and H-3 and of the deep parameters in Table H-4. The calibration is presented in Table H-5, the priors for those parameters that are not set using information from the DSGE models can be found in Table H-6, whereas data used in estimation are listed in H-7.

Appendix H.1. Parameter and Variable Definitions

Table H-2: Endogenous Variables.

Variable	Description
π	inflation
$\tilde{y}$	output gap
y^n	natural output
y	output
r^n	natural interest rate
r^r	real interest rate
$\hat{i}$	nominal interest rate
n	hours worked
$\hat{h}$	heterogeneity index
$\hat{\gamma}$	consumption gap
ν	AR(1) monetary policy shock
a	AR(1) technology shock
z	AR(1) preference shock
$r^{r,ann}$	annualized real interest rate
i^{ann}	annualized nominal interest rate
$r^{n,ann}$	annualized natural interest rate
π^{ann}	annualized inflation rate

Table H-3: Innovations to Exogenous Shocks.

Variable	Description
ε_v	monetary policy shock
ε_a	technology shock
ε_z	cost-push shock

Table H-4: Deep Parameters.

Variable	Description
β	discount factor
λ	fraction of constrained households
τ	profit transfers
δ	fraction illiquid to total assets
σ	log utility
φ	unitary Frisch elasticity
Ξ	price adjustment cost
ε_p	demand elasticity
ϕ_π	inflation feedback Taylor Rule
ϕ_y	output feedback Taylor Rule
ρ_a	autocorrelation exogenous monetary policy component
ρ_a	autocorrelation exogenous technology process
ρ_a	autocorrelation exogenous cost-push process
σ_v	standard deviation, innovation to exogenous monetary policy component
σ_a	standard deviation, innovation to exogenous technology process
σ_z	standard deviation, innovation to exogenous cost-push process

Table H-5: Priors $p(\theta)$.

Parameter	Distribution	Mean	Standard Deviation	Lower Bound	Upper Bound
β	fixed	0.9745	0	-	-
λ	Gaussian	0.2	0.1	0.1	0.3
τ	fixed	1	0	-	-
δ	fixed	0.92	0	-	-
σ	Gaussian	2.00	0.37	0.95	-
φ	Gaussian	1.00	0.50	-	-
θ	Beta	0.50	0.10	-	-
ε_p	fixed	10	0	-	-
ϕ_π	Gaussian	1.50	0.25	1.01	-
ϕ_y	Gaussian	0.125	0.20	0.00	-
ρ_v	Gaussian	0.50	0.20	0.30	0.98
ρ_a	Gaussian	0.70	0.30	0.30	0.98
ρ_z	Gaussian	0.70	0.30	0.30	0.98
σ_v	Uniform	-	-	0.10	0.50
σ_a	Uniform	-	-	0.20	0.80
σ_z	Uniform	-	-	0.20	0.80

NOTE: All parameters are common across model except for λ, which is set to 0 in the RANK model. The distribution, mean and standard deviation represent the unconstrained ditribution, which are cosntrained further by the bounds in the last two columns (- denotes no constraint).

Appendix H.2. Model Equations

$$\pi_t = \beta\, \mathbb{E}_t\{\pi_{t+1}\} + \kappa\, \tilde{y}_t + z_t$$

$$\text{where } \kappa = \omega\, (\sigma + \varphi)\,,\ \omega = \frac{\varepsilon_p}{\Xi\, \mathscr{M}_t^p} \text{ and } \mathscr{M}^p = \frac{\varepsilon_p}{\varepsilon_p - 1}$$

$$\tilde{y}_t = \tilde{y}_{t+1} - \frac{1}{\sigma\, (1 - \Phi)}\, \left(\hat{i}_t - \pi_{t+1} - r_t^n\right)$$

$$\text{where } \Phi = \lambda\, \frac{(\sigma + \varphi)\, \Psi}{1 - \lambda\, \gamma} \text{ and } \Psi = \frac{(1 - \lambda)\, (1 - \delta\, (1 - \tau))}{\left(1 - \lambda + \left(\mathscr{M}_t^p - 1\right)\, (1 - \lambda\, \delta\, (1 - \tau))\right)^2}$$

$$\hat{i}_t = \pi_{t-1}\, \phi_\pi + \tilde{y}_{t-1}\, \phi_y + \nu_t$$

$$r_t^n = -\sigma\, (1 - \rho_a)\, \psi_{ya}^n\, a_t$$

$$r_t^r = \hat{i}_t - \pi_{t+1}$$

$$y_t^n = \psi_{ya}^n\, a_t, \text{ where } \psi_{ya}^n = \frac{1 + \varphi}{\sigma + \varphi}$$

$$\tilde{y}_t = y_t - y_t^n$$

$$\hat{\gamma}_t = -\,(\sigma + \varphi)\, \Psi\, \tilde{y}_t$$

$$\hat{h}_t = \Phi\, \tilde{y}_t$$

$$\nu_t = \rho_a\, \nu_{t-1} - \varepsilon_t^\nu$$

$$a_t = \rho_a\, a_{t-1} + \varepsilon_t^a$$

$$z_t = \rho_a\, z_{t-1} + \varepsilon_t^z$$

$$y_t = a_t + n_t$$

$$i_t^{ann} = 4\hat{i}_t,\ r_t^{r,ann} = 4 r_t^r,\ r_t^{n,ann} = 4 r_t^n,\ \pi_t^{ann} = 4\pi_t$$

Appendix H.3. Priors for RANK/TANK Application

Table H-6 lists the priors used in our empirical application for those parameters that are not set using information from the DSGE models. The mean of an inverse gamma distribution is determined by a scale parameter *scale* and a shape parameter *shape* such that its mean (if it exists) is equal to $\frac{scale}{shape-1}$ and the variance (if it exists) is given by $\frac{scale^2}{(shape-1)^2(shape-2)}$.

Table H-6: Additional Priors.

Parameter	Distribution	Mean	Information on Standard Deviation
μ	Gaussian	0	1 for free elements, 0 else
$\Sigma_{u,i}$	Inverse Gamma $\forall i$	0.2^2	∞ (scale parameter of 2)
$\Sigma_{w,i}$	Inverse Gamma $\forall i$	0.5^2 for GDP, 0.2^2 else	shape parameter is set to $T/2$
$\tilde{X}_0$	Gaussian	0	10
μ^z	Gaussian	0.25 for GDP, 0 else	0.50 for GDP, 0 else

Appendix H.4. Data for RANK/TANK Application

In Table H-7 we report the data used in the RANK/TANK application.

Table H-7: Data Information for RANK/TANK Application.

Variable	Sample	Transformation
GDP	1970:Q1-2019:Q4	seasonally adjusted, per-capita, then log
GDI	1970:Q1-2019:Q4	seasonally adjusted, per-capita, then log
CPI	1970:Q1-2019:Q4	seasonally adjusted, log
PCE price index	1970:Q1-2019:Q4	seasonally adjusted, log
Three month treasury bill rate	1970:Q1-2019:Q4	average over the quarter
Federal Funds rate	1970:Q1-2019:Q4	average over the quarter
Wu-Xia shadow rate	1990:Q1-2019:Q4	average over the quarter

NOTE: The data is from the St. Louis Fed's FRED database, except for the Wu-Xia shadow rate, which is downloaded from https://www.atlantafed.org/cqer/research/wu-xia-shadow-federal-funds-rate.

Appendix H.5. Further Results for the RANK/TANK Application

To highlight how our approach can be used to estimate a cyclical component informed by economic theories, Figure H-2 plots the median and 5th as well as 95th percentiles for the cyclical component of real GDP (i.e. the posterior estimates of one element of X_t) in the left panel (NBER recessions are depicted using gray bars). We can see that the estimated cyclical component tends to decline during recessions, as predicted by our theories.

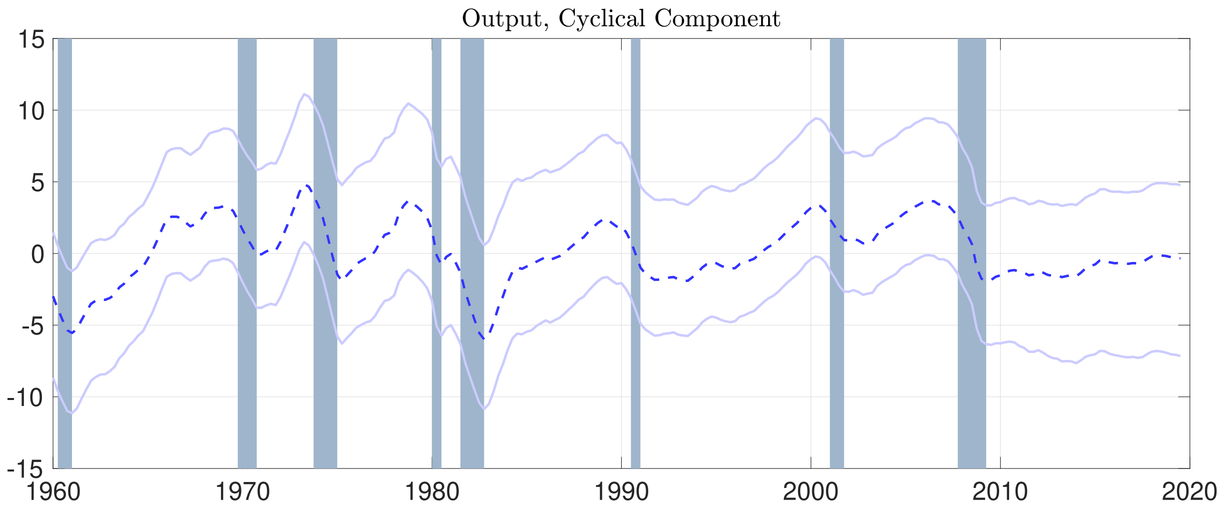

Figure H-2: Estimated Cyclical Component for (Log) Per-Capita Real GDP.

To understand what helps us discriminate between these two models, it is useful to dig deeper into the structure of the models. It turns out that in the setup of Debortoli and Galí (2018), heterogeneity only appears in the IS equation (i.e. the consumption Euler equation). Both the RANK and the TANK model share the same IS equation:

$$y_t = E_t y_{t+1} - \frac{1}{\sigma\left(1 - \Phi(\lambda)\right)}\left(i_t - E_t \pi_{t+1} - r_t^n\right), \tag{H-1}$$

where y_t is the output gap, π_t is inflation, i_t is the nominal interest rate, and r_t^n is the natural real rate in the model (all measured in deviations from steady state). The key object is $\Phi(\lambda)$, which is a composite parameter that depends, among other things, on the fraction of restricted households λ.

Figure H-3 shows the QQ plot of $\frac{1}{\sigma(1-\Phi)}$ based on the prior distributions $f_i(\theta_i)$.[30] This expression can be interpreted as the interest rate sensitivity of the output gap (holding expectations constant). Heterogeneity makes the distribution of this expression shift to the right (the TANK model quantiles are on the y-axis). It is this higher interest rate sensitivity that leads to the improved fit of the TANK-based VAR.

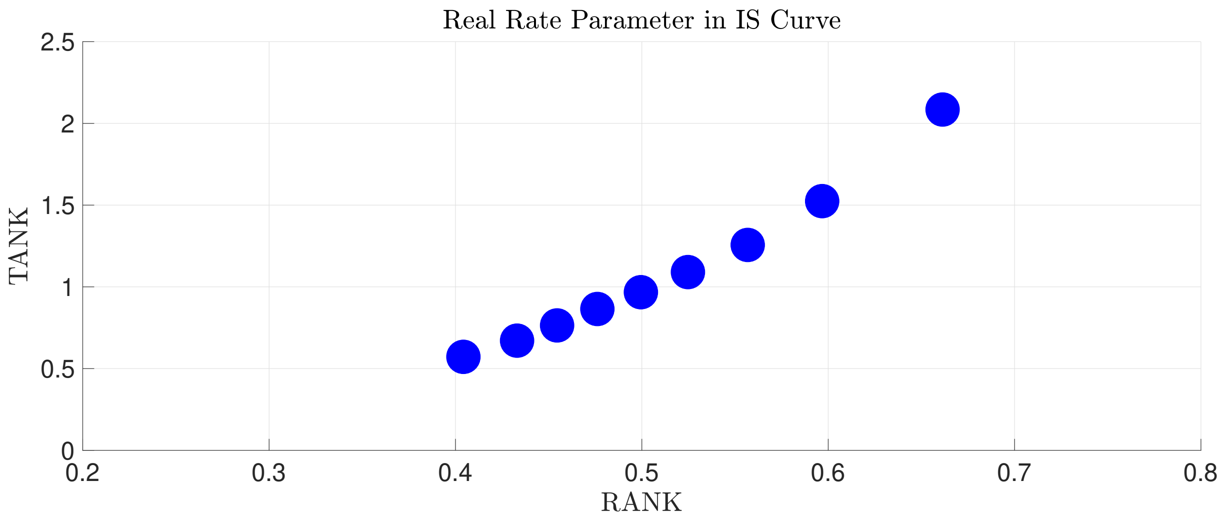

Figure H-3: QQ Plot for $\frac{1}{\sigma(1-\Phi(\lambda))}$ in TANK and $\frac{1}{\sigma}$ in RANK, from 10th to 90th Percentile.

Previous papers (Amisano and Geweke, 2017, Del Negro et al., 2016, and Waggoner and Zha, 2012, for example) have emphasized that in environments with misspecified models the associated weights (or probabilities) can vary substantially over time. To assess this, we also estimate the models weights recursively, starting the sample at the same datapoint as the total sample, but ending it after ten years. We then incrementally add 1 year at a time to the sample until we reach the end of our total sample. Figure H-4 plots the estimated median as well as the 5th and 95th percentile bands for the model weights.[31] While there is some fluctuation (and in particular a substantial increase in the uncertainty around 1990), the TANK model remains the preferred model throughout the sample.

[30] A QQ plot is a scatter plot of the quantiles of two different distributions. For example, the leftmost dot in the plot represents the 10th quantile of both distributions.

[31] For the sake of comparison, we use the same VAR prior (which is the one we used for the full sample) for all samples.

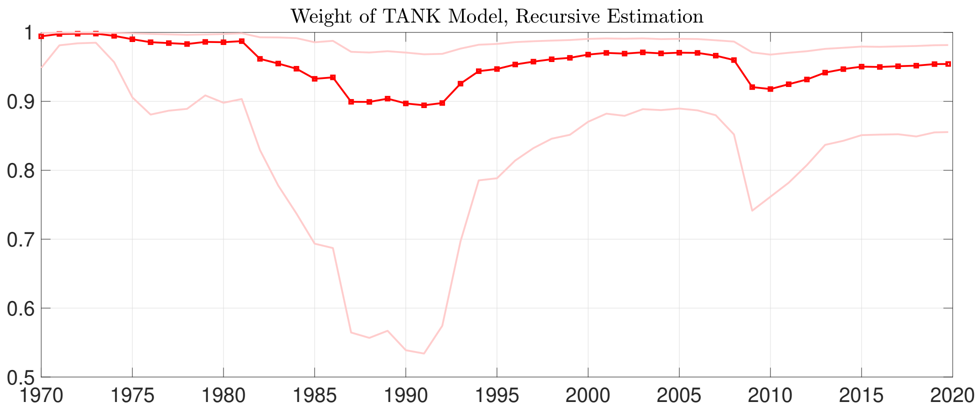

Figure H-4: Recursive Model Weights (5th, 50th, and 95th Percentiles).

Our approach is different from aforementioned pooling approaches such as Amisano and Geweke (2017) or Del Negro et al. (2016) when those approaches directly pool dynamic equilibrium models. In such pooling approaches, the one-step ahead forecast density has to be a weighted average of the forecast densities from the individual equilibrium models. Our approach instead delivers a forecast density that generally is not a weighted average of the densities coming directly from the equilibrium models, and thus has an advantage if the individual equilibrium models are severely misspecified.

Appendix H.6. Forecasting Performance in the RANK/TANK Application

We next study the forecasting performance of our approach. To get a sense of how well our model performs, we study the relative root mean squared error (RMSE) for our six observables relative to a benchmark forecasting model. We recursively add data to both models and re-estimate every year (the recursive estimation of our model also delivered Figure H-4). As this benchmark forecasting model we choose an ARIMA(p,d,q) model where p denotes the AR lag length, q denotes the MA lag length, and d denotes the order of integration. We select the model specification using the Akaike information criterion (AIC) to maximize forecasting performance. For each recursive sample, we first conduct ADF tests to determine the order of integration. The test specification includes a constant and the lag length is set based on the AIC. Then, we determine p and q also based on the AIC where the maximum p and q are set to 8. Table H-8 shows the relative RMSE, where a number larger than 1 implies a better performance of the ARIMA model. Two findings stand out: (i) similar to Del Negro and Schorfheide (2004), our approach does better at medium horizons compared to very short horizons, and (ii) our approach performs better for the nominal variables in our sample. In order to improve forecasting performance for GDP and GDI, we could modify the law of motion for z_t to allow for richer trend dynamics (see for example Canova, 2014).

Table H-8: Relative RMSE (of VAR/ARIMA) by Forecast Horizon h (in Quarters).

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
GDP	1.4	1.2	1.2	1.1	1.1	1.1	1.2	1.2
GDI	1.2	1.1	1.1	1.1	1.1	1.1	1.2	1.2
CPI Inflation	1.0	0.9	0.8	0.9	0.9	0.9	0.8	1.0
PCE Inflation	1.0	1.0	1.0	1.0	1.0	0.9	0.9	1.0
3-Month Treasury Bill	1.2	0.8	0.8	0.9	1.0	1.1	1.1	1.0
Federal Funds Rate	1.2	1.1	0.8	0.8	0.9	0.9	0.9	0.9

Appendix I. More on the Relationship Between Our Approach and Standard Bayesian Model Probabilities

It is well known that standard Bayesian model probabilities include an overfitting penalty, as is evident, for example, from the form of the well-known Bayesian Information Criterion, which approximates the Bayes factor in one particular situation. In standard marginal likelihood comparisons, overfitting gets penalized because having more parameters can lead to substantially worse fit for some parameter values, which a marginal likelihood comparison will acknowledge since it is the integral of the product of prior and likelihood over the entire parameter space. In our approach something similar can happen, although the effects are likely muted: Our approach builds on marginal likelihoods for VARs with different priors, but the same number of parameters. Priors for the VARs associated with theories (or DSGE models) with more parameters might put more prior mass on regions of the parameter space that do not fit the data as well because these VAR priors are informed by the priors for the DSGE models. So parameter combinations that do not fit the data as well do still enter the marginal likelihood comparison for the VAR structure via the VAR priors.