Economic Theories and Macroeconomic Reality

Francesca Loria — *Federal Reserve Board*
Christian Matthes — *Indiana University*
Mu-Chun Wang — *Deutsche Bundesbank*

January 6, 2021

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 show that household heterogeneity in New Keynesian theories greatly helps to fit data on interest rates, output, and inflation.

Keywords: Bayesian Inference, Misspecification, Heterogeneity, VAR, DSGE
JEL Classification: C32, C50, E30

We thank Gianni Amsiano, Christiane Baumeister, Joshua Chan, Siddhartha Chib, Thorsten Drautzburg, Thomas Drechsel, Ed Herbst, Dimitris Korobilis, Elmar Mertens, 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, and the Drautzburg-Nason workshop for their useful comments and suggestions on earlier versions of the paper.

Disclaimer: 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.

"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.

We propose to use a suite of economic theories / equilibrium models to simulate data. Based on these simulated datasets from the various models we build a mixture-prior for coefficients of a Vector Autoregression (VAR, Sims (1980)) in possibly mis-measured variables (where we explicitly model the mis-measurement). Each mixture component is informed by one theory. Our algorithm then automatically generates information about which theory / implied prior best fits the data by updating the weights associated with each mixture component / theory. Because we allow deviations of the posterior of our coefficients from either 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.

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.[1] Our interest is different from theirs: we want to distinguish between possibly non-nested models

[1] 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).

that could not or should not be taken to the data directly.

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 be 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 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 (2018) 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 (2018) 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).

One contribution of our paper is to propose 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 (2019) 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. Nonetheless, the mixture priors we introduce for VARs 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).

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

We then use our approach to ask whether modeling heterogeneity across households has a substantial effect for the fit of New Keynesian models. In particular, we study per capita output, inflation, and nominal interest rates, which are common observables in the study of New Keynesian models (see for example Lubik and Schorfheide (2004)). 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 built on Debortoli and Galí (2017), who show that a two agent New Keynesian model can already approximate a richer heterogeneous model in terms of many aggregate implications.[2] We thus contrast a version of Debortoli and Galí (2017)'s representative agent model with a version of their two agent model where some agents are constrained in that they cannot participate in asset markets (they are "hand-to-mouth" consumers) and find that even a model with this relatively stark form of heterogeneity is much preferred by the data to its representative agent counterpart. The tractability of this two-agent model allows us to pinpoint the channel through which household heterogeneity influences aggregate dynamics and improves aggregate fit.

The paper is structured as follows: The next section introduced the econometric framework, section 4 contains our empirical application, and sections 6 and 7 provide an illustration and a set of Monte Carlo studies of our methodology to highlight (i) differences between our approach and the standard approach that directly computes the marginal likelihood of each structural model, and (ii) how our approach *averages* across models with non-degenerate

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

weights if the data calls for this.

2 Econometric Framework

Our goal is to build a mixture prior for a VAR of unobserved state variables where each mixture component is informed by one specific economic theory (or equilibrium model). We first give a broad overview of our approach, summarized in Figure 1, before going into the details of each step.

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 $p_i(\theta_i)$ for the vector of deep parameters θ_i of model i and the associated distribution of the data conditional on a specific parameter vector $p_i(X_i^t|\theta_i)$ (Gelman *et al.*, 2013). For each of the K models, we only require that we can simulate from $p_i(\theta_i)$ and $p_i(X_i^t|\theta_i)$.[3] The simulation from the models constitutes one block or module of our algorithm.

Given simulations from each model, we construct a prior from a conjugate family for a VAR on a common set of variables $X^t \in X_i^t$, $\forall i$. 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).[4] This mapping from simulations to VAR priors is the second block or module of our approach.

Next, 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 to the fit (marginal likelihood) of the VAR model with each mixture com-

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

[4] 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.

ponent 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, 2019). 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).

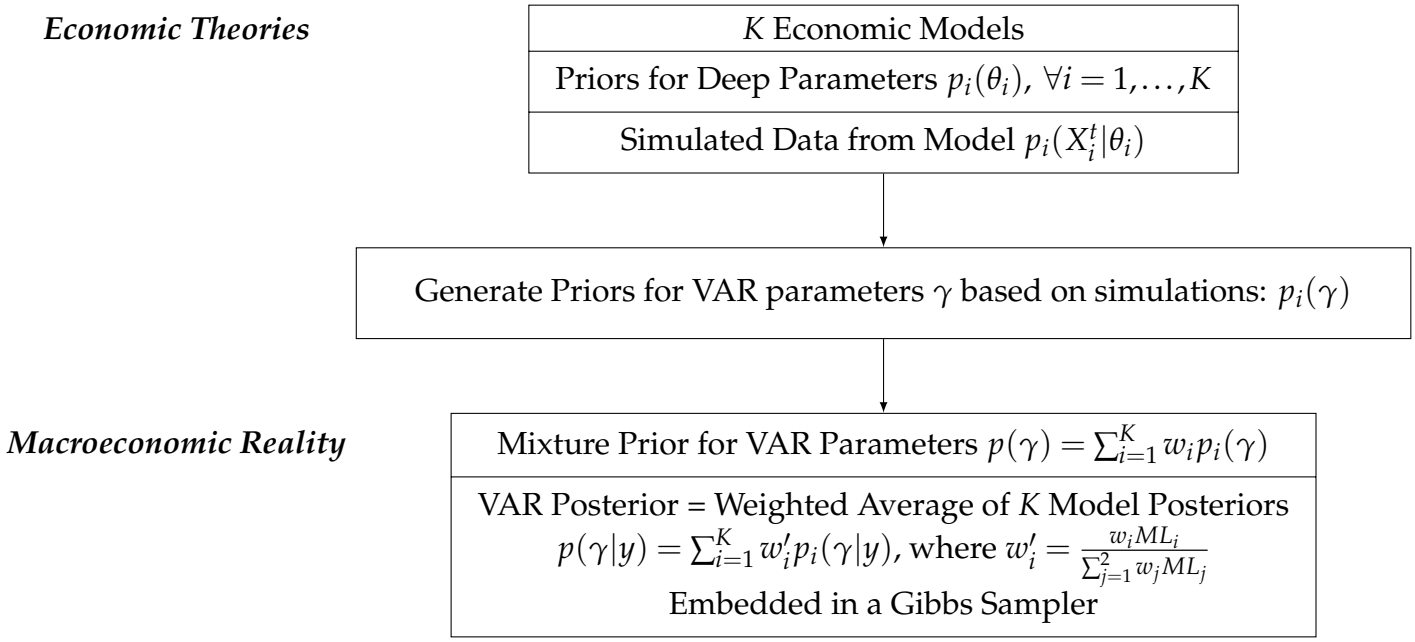

Figure 1: From Economic Theories to Macroeconomic Reality.

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 $K > 2$ 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 details, 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 $\mathcal{F}$ of sampling distributions $p(y|\gamma)$ and $\mathcal{P}$ a class of prior distributions. then the class $\mathcal{P}$ is* conjugate *for $\mathcal{F}$ if*

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

A class $\mathcal{P}$ 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 by y a generic vector of data.

To highlight the flexibility that a mixture prior gives us, Figure 2 (inspired by a similar figure in Rossi (2014)) plots four mixtures of two Gaussian distributions. Even with two Gaussian components, these examples show patterns researchers might find useful when setting priors: fat tails, skewness, multimodality, and putting more mass around the mode.

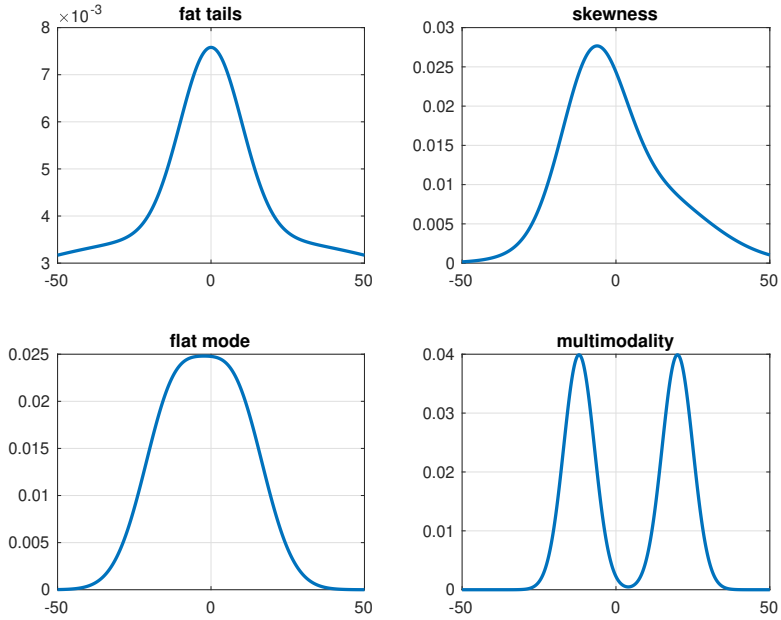

Figure 2: Various two-component Gaussian mixtures.

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:[5]

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

where

$$w_i' = \frac{w_i ML_i}{\sum_{j=1}^{2} w_j ML_j}, \ \forall i = 1,2. \tag{2.1}$$

$$ML_i = \int p_i(\gamma) p(y|\gamma) d\gamma \tag{2.2}$$

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.

2.1 Our Time Series Model - *Macroeconomic Reality*

With this general result in hand, we turn to the specific time series model that we use. We embed the VAR described above in 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. Should our models match inflation based on the CPI, PCE, or the 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 the 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.[6] We thus follow Canova (2014) and allow for unobserved components that are persistent and not related to the economic theories we consider. To be specific,

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

[6] A telltale sign of this in macroeconomics is that data and model outputs are often filtered before being compared to each other.

our time series model is:[7]

$$Y_t = \mu + AX_t + Bz_t + u_t, \tag{2.3}$$

$$X_t = \sum_{j=1}^{J} C_j X_{t-j} + \varepsilon_t, \tag{2.4}$$

$$z_t = \mu^z + z_{t-1} + w_t \tag{2.5}$$

where $u_t \overset{iid}{\sim} N(0, \Sigma_u)$ is a vector of measurement errors with a diagonal covariance matrix Σ_u, $\varepsilon_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon)$ and $w_t \sim N(0, \Sigma_w)$. 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. 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.[8] 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.[9] 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, 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 (2.4) 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.

Estimation via Gibbs Sampling With a mixture prior for the VAR coefficients $\{C_j\}_{j=1}^J$ and Σ_ε in hand (the construction of which we describe below), we can approximate the posterior via Gibbs sampling in three steps or blocks. We focus here on an overview of the algorithm (details can be found in Appendix A). Remember that a Gibbs sampler draws parameters for a given block conditional on the parameters of all other blocks. One sweep through all

[7] We do not mean to imply that this time series might not be misspecified along some dimensions. We think of it as a good descriptions of many features of aggregate time series (more so than the economic theories we consider). One could enrich our model 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.

[8] 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.

[9] 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 empirical application do not have that feature.

blocks then delivers one new set of draws. The first block draws 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)). The second block consists of the parameters for the measurement equation μ, A, and Σ_u.[10] 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 the each diagonal element of Σ_u, which allows us to draw from the conditional posterior for those variables in closed form. Finally, the VAR coefficients $\beta = vec([\{C_j\}_{j=1}^J])$ and Σ_ε are drawn according 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 β and Σ_ε 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), and (iii) the prior recently proposed by Chan (2019) that breaks the cross-equation correlation structure imposed by the first prior. We use (iii) as our benchmark as it allows for a more flexible prior for β 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 straightforwards. We use prior (ii) for robustness checks - we find prior (i) too restrictive for actual empirical use. Another advantage of the prior structure introduced in Chan (2019) 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:

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

where $\mathbf{F}_i$ is the mapping describing the law of motion of the state variables and $\epsilon_{t,i}$ are the structural shocks in model i at period t. 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 questions. 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 below). For our

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

algorithm we need N draws from $p_i(\theta_i)$. For each of these draws, we simulate a sample of length T, which we discuss in the next section. Note that while solving models nonlinearly can be time consuming, this step of the algorithm can generally be carried out in parallel. 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.[11]

2.3 Generating VAR Priors from Simulated Data

Mapping Economic Theories Into Macroeconomic Reality

We have K models with parameter vector θ_i and associated prior $p_i(\theta_i)$ we want to derive VAR priors from.[12] For a given n-dimensional vector of observables Y_t, we need a prior for the VAR coefficients β and the residual covariance matrix Σ_u. We use a simulation-based approach to set our priors. This not only easily generalizes to non-linear DSGE models but also allows us to take parameter uncertainty into account easily.

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}. 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 β, 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 B).

We pick the number of simulated data sets R to be 2000 in our empirical application. We generally recommend to increase the number of simulations until the corresponding prior does not change any more. 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 actual data.[13] 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.

[11] We find it useful to assume that the dimension of $\epsilon_{i,t}$ 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.

[12] 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).

[13] 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.

Finally, 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.

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:

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

where $x_{1,t}$ and $x_{2,t}$ are demeaned variables simulated from an equilibrium model. We know that asymptotically $\hat{\beta} = \frac{cov(x_{1,t},x_{2,t})}{var(x_{2,t})}$, where *cov* and *var* are population moments. Under well known conditions, the regression coefficient on long simulations from the model will approach $\hat{\beta}$. 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 $\hat{\beta}$, 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.[14]

3 A Possible Extension: Impulse Responses and Data-Driven Averaging Over Identification Restrictions

Consider a situation where for each model i we have a function $\mathcal{M}_i(\gamma)$, where γ is the vector of VAR coefficients. This function 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 $\mathcal{M}_i$ functions, namely that the associated identification

[14] 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).

restrictions either lead to set-identification or exact identification.[15] 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 $\mathcal{M}_i$ function for each model (for example, the choice of what sign restrictions are imposed and for what horizons).

Remember the structure of our Gibbs sampler: we draw a model indicator according to the implied model probabilities based on the marginal likelihoods of the VARs, and then conditional on that indicator we draw VAR parameters. We can then add an additional step right after that generates a draw from $\mathcal{M}_i(\gamma)$ (we could also do this ex-post after the reduced-form estimation). We can then 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 would take into account uncertainty about identification schemes. If one model gets almost all the probability weight, we will basically only use identification restrictions from that model.

We can also generalize the approach and allow $n_i = 1, ..., N_i$ identification schemes *per model*, each associated with a prior probability $p_i^{n_i}$. 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.[16]

4 Does Heterogeneity Matter for Aggregate Behavior?

In recent work, Debortoli and Galí (2017) 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 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.

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

[16] 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.

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 (2017) also show that TANK models approximate reasonably well the aggregate output dynamics of a canonical HANK model in response to monetary and non-monetary shocks.
We use versions of these models to simulate data on log output, hours worked, real interest rate and productivity.[17] The model equations and the calibration of the model are presented in Appendix C. The priors for those parameters that are not informed by the DSGE model are described in detail in Appendix D. All DSGE priors (the $p_i(\theta_i)$ 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 full descriptions of the models (including the full description of the DSGE priors) can be found in Appendix C. 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.
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 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

[17] The only differences relative to Debortoli and Galí (2017) 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.

trend inflation described by, for example, Cogley and Sbordone (2008)).

Table 1: Posterior Mean of Model Weights, RANK vs. TANK. Benchmark Specification and Two 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

In table 2, we show the resulting model probabilities. 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, and (ii) we use a common Normal-inverse Wishart Prior instead of the Chan (2019) prior. As highlighted in Table 2, 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 and modify T^{final}, the length of the simulated data series that are used to inform our prior. We set this sample size to half the actual sample size in this robustness check. 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.

Table 2: Posterior Mean of Model Weights, RANK vs. TANK. Additional Robustness Checks.

	RANK	TANK
T^{final} equal to half the actual sample size	0.01	0.99
VAR(4)	0.13	0.87

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 3 plots the estimated median as well as the 5th and 95th percentile bands for the model weights.[18] 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. 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.

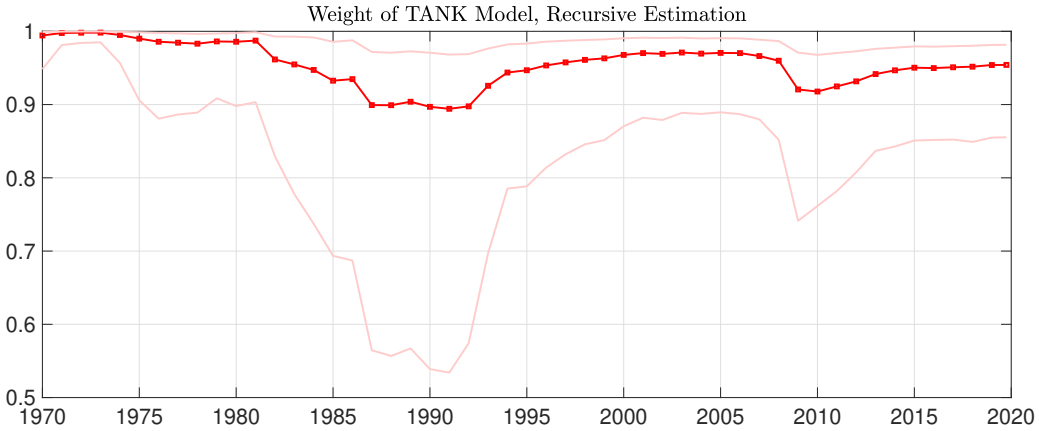

Figure 3: Recursive Model Weights (5th, 50th, and 95th percentiles).

To highlight how our approach can be used to estimate a cyclical component informed by economic theories, Figure 4 plots the median and 5th as well as 95th percentiles for the cyclical component for real GDP 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.

[18] For the sake of comparison, we keep use the VAR prior from the full sample for all samples.

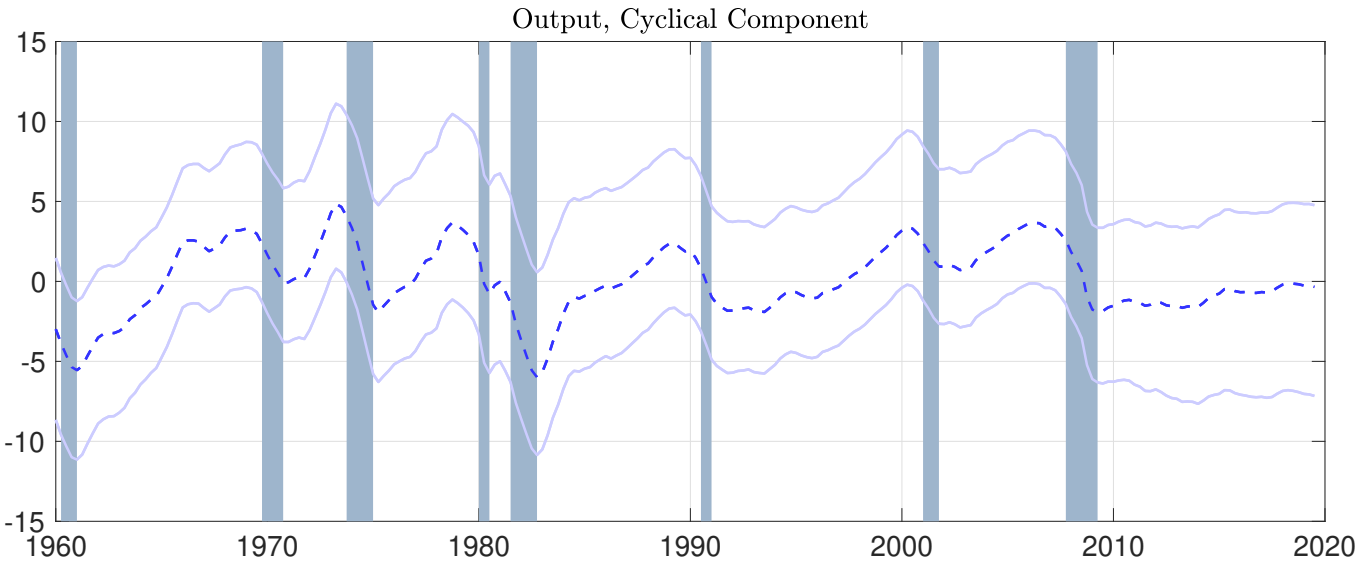

Figure 4: Estimated Cyclical Component for (log) per-capita real GDP.

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 (2017), 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:

$$y_t = E_t y_{t+1} - \frac{1}{\sigma\left(1-\Phi(\lambda)\right)}\left(i_t - E_t\pi_{t+1} - r_t^n\right), \tag{4.1}$$

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 5 shows the QQ plot for the distribution of $\frac{1}{\sigma(1-\Phi)}$.[19] 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.

5 VAR Forecasting Performance

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

[19] 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.

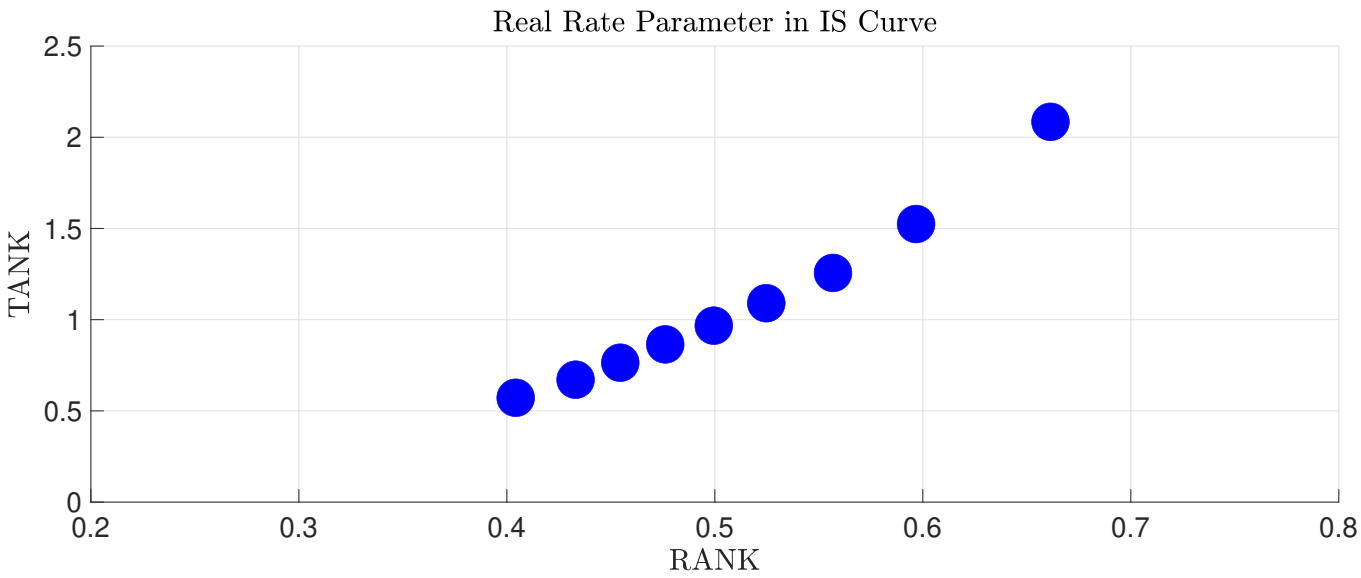

Figure 5: QQ plot for $\frac{1}{\sigma(1-\Phi)}$, from 10th to 90th percentile.

and re-estimate every year (the recursive estimation of our model also delivered Figure 3). 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 3 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)).

We do want to further analyze what a 40 percent increase in RMSE means (as is the case for GDP at the one-quarter horizon). In particular, does this mean that visual inspection will make it clear that our approach forecasts substantially worse? To do this, Figure 6 plots the data for GDP and GDI as well as the one quarter ahead forecasts from both our approach and the ARIMA model. The forecasts are aligned with the data they aim to forecast. Visually the forecasts are basically impossible to distinguish. We think our approach thus broadly captures the dynamics of real output, even if the ARIMA model outperforms it at short horizons.

Table 3: Relative RMSE (RMSE of VAR/RMSE of ARIMA) as a function of the 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	1.2	1.1	0.8	0.8	0.9	0.9	0.9	0.9

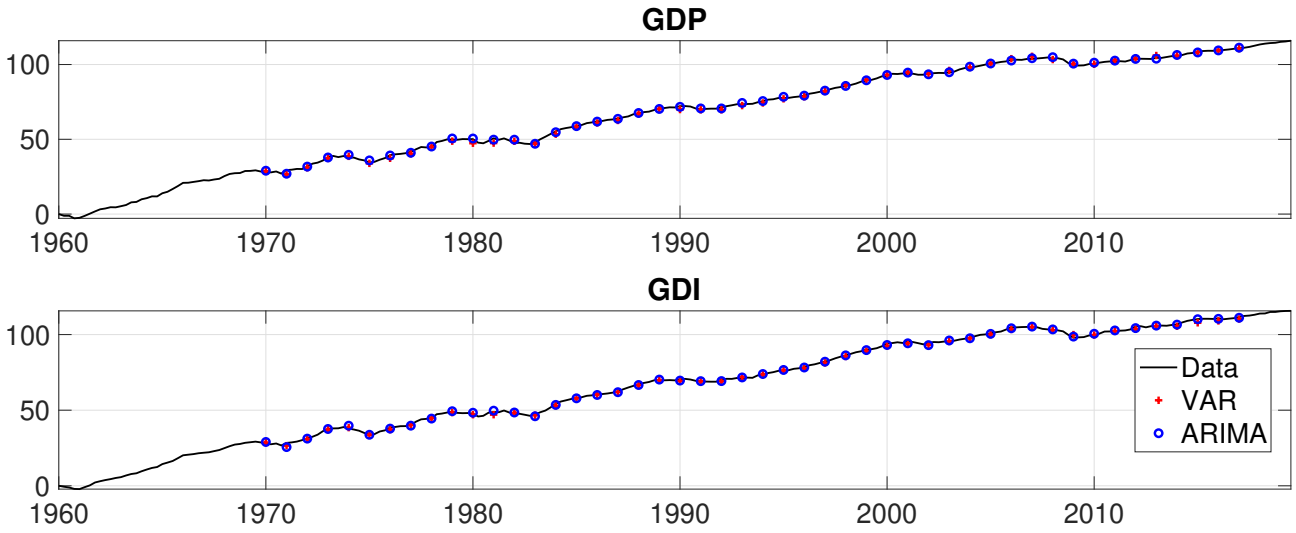

Figure 6: Recursive one-quarter ahead forecasts for (log) real per-capita output.

6 An Illustration

To get a sense of what our approach is doing 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 prior mixture block of the Gibbs sampler - this allows us to plot priors and posteriors easily.

In Figure 7, 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 (2.2). While component 2 is favored in this example, compo-

nent 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.

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.[20]

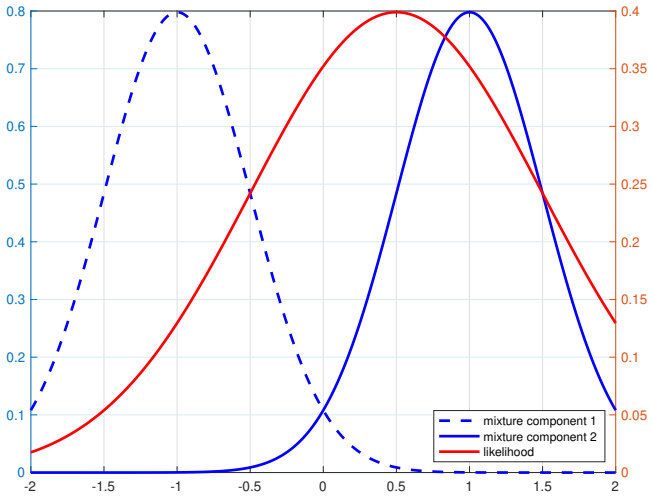

Figure 7: Components of Mixture Prior and Posterior.

Let's turn to a scenario where the likelihood has less variance, as depicted in Figure 8. What we can see 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

[20] By "likelihood function for X_t" we mean the density of X_t conditional on parameters, as depicted for example, in Figure 7.

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.

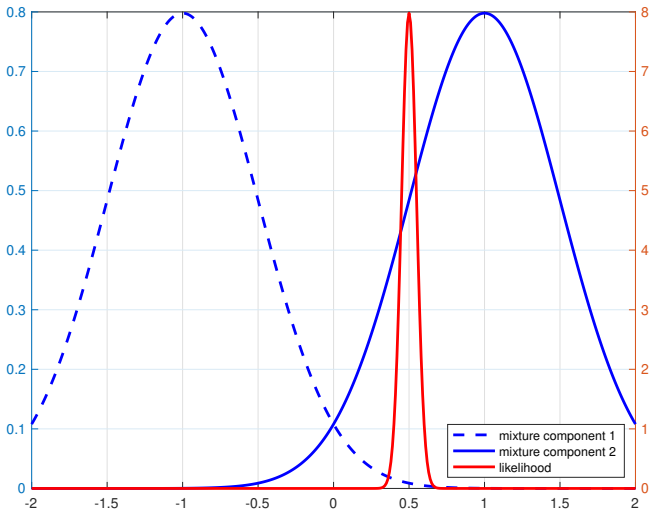

Figure 8: Components of Mixture Prior and Posterior, Tighter Posterior.

7 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 contrasts our approach and the standard marginal likelihood-based approach based directly on the likelihood of the individual models. Then we highlight that our approach automatically provides model averaging if the various models all fit the data well and further analyze the role that the iid measurement error plays in our setup.

7.1 Our Approach vs. a Standard Approach

Throughout the first three examples, we want to discriminate between the following two time series models for the bivariate vector of observations $[x_t \; y_t]'$:

Model 1

$$y_t = 0.95 y_{t-1} + e_t, \tag{7.1}$$

$$x_t = y_t + v_t, \tag{7.2}$$

where $e_t \overset{iid}{\sim} N(0,1)$ *and* $v_t \overset{iid}{\sim} N(0,0.45^2)$.

Model 2

$$y_t = 0.95y_{t-1} + e_t, \tag{7.3}$$

$$x_t = 0.95x_{t-1} + u_t, \tag{7.4}$$

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

Note that both models generate variables with the same persistence, but the comovement patterns are very different. All parameters are assumed to be known (in the language of Section 2, $p_i(\theta_i)$ is degenerate for $i = 1,2$). [21] We want to contrast two ways of figuring out which model fits the data better (we will discuss results for various data-generating processes below). First, we can follow our approach where we compute the average model probability across draws for each sample. Alternatively, we can compute the marginal likelihood for each model. Since all parameters are assumed to be known, this amounts to just computing the value of the likelihood function at the known parameters. We will report the log of the likelihood ratio. First, let us consider the case where model 1 is the data-generating process. We compare approaches using 1000 Monte Carlo samples of length 240 each. To set the priors, we simulate for each Monte Carlo sample 2000 samples of the same length as the sample used in estimation. Table 4 shows the average model probability (across draws and samples) and the log of the likelihood ratio as well as the associated standard deviation.

Table 4: Evidence for Model 1 when Model 1 is the Data-Generating Process.

	Model Probabilities	Log Odds Ratio
Mean	1.00	225.54
Standard Deviation	0.00	29.39

Before going into the interpretation, we repeat the same exercise, but with model 2 as the data-generating process - the results are depicted in table 5.

[21] For parsimony, we do not consider random walk components in our time series models in this section. The prior model weights are set at 0.5. In this example, the exact choice of the natural conjugate prior is inconsequential - the results here are obtained using the Normal-inverse Wishart prior. The parameters in the measurement equations are also assumed to be known and fixed in all exercises in this section. We achieve this by using priors that are tightly centered at the true parameter values.

Table 5: Evidence for Model 1 when Model 2 is the Data-Generating Process.

	Model Probabilities	Log Odds Ratio
Mean	0.00	-2358.53
Standard Deviation	0.00	1036.04

The evidence is clear: When our set of models contains the truth, we can clearly pick out the truth in realistic sample sizes.[22] Furthermore, there is no substantial difference between model choice (or model averaging) between our approach and an approach based directly on the likelihood function of the models.

Next, we turn to a more interesting and realistic question: What happens if both models are misspecified? As a data-generating process we now use the following model:

Model 3

$$y_t = 0.95y_{t-1} + e_t, \tag{7.5}$$

$$x_t = y_t + w_t, \tag{7.6}$$

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

This model looks deceptively like model 1, but the noise term w_t has a larger variance. Nonetheless, it features the same low frequency comovement between x_t and y_t that is a hallmark of model 1. As such, we think many economists would hope/expect that model 1 would better fit the data generated by model 3. Table 6 shows that this hope is on average confirmed by our procedure with an average model probability of 0.83. The likelihood-ratio paints a different picture, however. Here the average is substantially negative, favoring model 2. We also report the fraction of Monte Carlo samples that prefer model 1 in the sense that either the model probability for model 1 is larger than 0.5 or the log-likelihood ratio is larger than 0.

When all models are misspecified, our approach emphasizes different features of model fit than a likelihood-based approach. Asymptotically, marginal likelihoods will be larger for a model the smaller the Kullback-Leibler distance to the true model is (Fernandez-Villaverde and Rubio-Ramirez, 2004). But that distance might be in part determined by features that economists might think are not as important to understanding the data-generating process, a point also made by Kocherlakota (2007).

[22] Even though the standard deviation of the log odds ratio seems large, all Monte Carlo samples yield a negative log odds ratio.

Table 6: Evidence for Model 1 when Model 3 is the Data-Generating Process.

	Model Probabilities	Log Odds Ratio
Mean	0.83	-52.41
Standard Deviation	0.35	45.04
Fraction of Samples Favoring Model 1	0.82	0.12

7.2 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 another 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. The two models we consider are:

Model 4 *Less Persistence*

$$y_t = 0.7y_{t-1} + e_t, \tag{7.7}$$

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

Model 5 *More Persistence*

$$y_t = 0.9y_{t-1} + e_t, \tag{7.8}$$

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). [23]

[23] In terms of specification, 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.

Table 7: 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 the middle of the sample	0.48	0.52

We can see 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.

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 iid. 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 two is more persistent, which the iid measurement error cannot mask.[24]

8 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, such as the ability to estimate cyclical components and trends, to produce reasonable forecasts of macroeconomic aggregates, and to compute impulse responses. Our approach enriches these standard approaches by allowing researchers to use various economic theories to inform the priors, and aid with identification of structural shocks.

In our empirical application, we show that even stylized theories of household heterogene-

[24] An autocorrelated measurement error would certainly alter this conclusion, but such a change in our setup, while technically feasible, goes against the spirit of our exercise – we contribute all persistent movements in the data to either cyclical movements or trends.

ity can help to fit standard macroeconomic data. Our approach could be used to further analyze issues concerning HANK models, for example the importance of countercyclical idiosyncratic risk in fitting aggregate data (Acharya and Dogra (2020)). .

References

ACHARYA, S. and DOGRA, K. (2020). Understanding HANK: Insights From a PRANK. *Econometrica*, **88** (3), 1113–1158.

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

ARUOBA, S. B., DIEBOLD, F. X., NALEWAIK, J., SCHORFHEIDE, F. and SONG, D. (2016). Improving GDP measurement: A Measurement-Error Perspective. *Journal of Econometrics*, **191** (2), 384–397.

BILBIIE, F. O. (2018). *Monetary Policy and Heterogeneity: An Analytical Framework*. CEPR Discussion Papers 12601, C.E.P.R. Discussion Papers.

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

BOX, G. and DRAPER, N. (1987). *Empirical Model-Building and Response Surfaces*. Wiley Series in Probability and Statistics, Wiley.

CANOVA, F. (1994). Statistical Inference in Calibrated Models. *Journal of Applied Econometrics*, **9** (S), 123–144.

— (2014). Bridging DSGE models and the raw data. *Journal of Monetary Economics*, **67** (C), 1–15.

— and MATTHES, C. (2018). *A Composite Likelihood Approach for Dynamic Structural Models*. Working Paper 18-12, Federal Reserve Bank of Richmond.

CHAN, J. C. C. (2019). *Asymmetric Conjugate Priors for Large Bayesian VARs*. CAMA Working Papers 2019-51, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

CHARI, V. V., KEHOE, P. J. and MCGRATTAN, E. R. (2007). Business cycle accounting. *Econometrica*, **75** (3), 781–836.

CHIB, S. and ERGASHEV, B. (2009). Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association*, **104** (488), 1324–1337.

— and TIWARI, R. C. (1991). Robust Bayes analysis in normal linear regression with an improper mixture prior. *Communications in Statistics*.

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

— and STARTZ, R. (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*, vol. 40A, Emerald Publishing Ltd, pp. 133–155.

DEBORTOLI, D. and GALÍ, J. (2017). *Monetary Policy with Heterogeneous Agents: Insights from TANK Models*. Economics Working Papers 1686, Department of Economics and Business, Universitat Pompeu Fabra.

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

DEL NEGRO, M., HASEGAWA, R. B. and SCHORFHEIDE, F. (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics*, **192** (2), 391–405.

DEL NEGRO, M. and SCHORFHEIDE, F. (2004). Priors from General Equilibrium Models for VARS. *International Economic Review*, **45** (2), 643–673.

DEL NEGRO, M. and SCHORFHEIDE, F. (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. and DRECHSEL, T. (2018). *Agnostic Structural Disturbances (ASDs): Detecting and Reducing Misspecification in Empirical Macroeconomic Models*. CEPR Discussion Papers 13145, C.E.P.R. Discussion Papers.

DURBIN, J. and KOOPMAN, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

FEDOROFF, A. *et al.* (2016). A geometric approach to eigenanalysis-with applications to structural stability.

FERNANDEZ-VILLAVERDE, J. and RUBIO-RAMIREZ, J. (2004). Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach. *Journal of Econometrics*, **123** (1), 153–187.

FESSLER, P. and KASY, M. (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., HARRISON, R. and THEODORIDIS, K. (2020). DSGE-based priors for BVARs and quasi-Bayesian DSGE estimation. *Econometrics and Statistics*, **16**, 1 – 27.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall.

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

— and AMISANO, G. (2011). Optimal prediction pools. *Journal of Econometrics*, **164** (1), 130 – 141, annals Issue on Forecasting.

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

—, — and — (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 University Press.

GREGORY, A. W. and SMITH, G. W. (1991). Calibration as testing: Inference in simulated macroeconomic models. *Journal of Business & Economic Statistics*, **9** (3), 297–303.

INGRAM, B. F. and WHITEMAN, C. H. (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., KUO, C.-H. and ROSSI, B. (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., MOLL, B. and VIOLANTE, G. L. (2018). Monetary policy according to hank. *American Economic Review*, **108** (3), 697–743.

KOCHERLAKOTA, N. R. (2007). Model Fit and Model Selection. *St. Louis Fed Review*, pp. 349–360.

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

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

LEE, P. M. (2012). *Bayesian Statistics - An Introduction*. Wiley.

LUBIK, T. A. and SCHORFHEIDE, F. (2004). Testing for Indeterminacy: An Application to U.S. Monetary Policy. *American Economic Review*, **94** (1), 190–217.

ROSSI, P. (2014). *Bayesian Non- and Semi-parametric Methods and Applications*. Princeton University Press.

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.

VILLANI, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics*, **24** (4), 630–650.

WAGGONER, D. F. and ZHA, T. (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 XIA, F. D. (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking*, **48** (2-3), 253–291.

A 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].

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 $\sigma^2_{w,i}$ conditioned on z_i^T from the inverse Gamma distribution $IG\left(\alpha_i^w, \beta_i^w\right)$ for $i = 1, ..., M$ with $\alpha_i^w = \alpha_i^{w,0} + \frac{T}{2}$ and $\beta_i^w = \beta_i^{w,0} + \frac{\sum\left(z_{i,t} - \mu_{z,i} - z_{i,t-1}\right)^2}{2}$, 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 $V^*_{z,i} = \frac{1}{\frac{1}{V^0_{z,i}} + \frac{T}{\sigma^2_{z,i}}}$ and $\mu^*_{z,i} = V^*_{z,i}\left(\frac{\mu^0_{z,i}}{V^0_{z,i}} + \frac{\sum\left(z_{i,t} - z_{i,t-1}\right)}{\sigma^2_{z,i}}\right)$, where $\mu^0_{z,i}$ and $V^0_{z,i}$ are prior hyperparameters of $N\left(\mu^0_{z,i}, V^0_{z,i}\right)$.

4. Define $\tilde{Y}^T = Y^T - \tilde{A}\tilde{X}_t$. Σ_u is diagonal. Draw each diagonal element $\sigma^2_{u,j}$ 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 $\beta_j^u = \beta_j^{u,0} + \frac{\sum\left(\tilde{y}_{j,t} - \mu_j\right)^2}{2}$, 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 $V_j^* = \frac{1}{\frac{1}{V_j^0} + \frac{T}{\sigma^2_{u,j}}}$ and $\mu_j^* = V_j^*\left(\frac{\mu_j^0}{V_j^0} + \frac{\sum \tilde{y}_{j,t}}{\sigma^2_{u,j}}\right)$, 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 (2019)). 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.

[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.

B 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. Several 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

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

   where Σ is $M \times M$ matrix of residual covariance, Φ 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 $\hat{\Phi}$ 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 *et al.* (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 than solve for

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

2. Our benchmark approach uses the asymmetric conjugate prior by Chan (2019). The essential assumption is that VAR coefficients are independent across equations. Using the notation from Chan (2019), 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 coef-

ficient 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 $1/\hat{\sigma}_i^2$.

C Debortoli and Galí (2017): RANK vs. TANK

The model is described by the following set of equations, where we provide the description of the endogenous and exogenous variables in Tables A-1 and A-2 and of the deep parameters in Table A-3. The calibration is presented in Table A-4.

C.1 Parameter and Variable Definitions

Table A-1: 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 A-2: Innovations to Exogenous Shocks.

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

Table A-3: 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_ν	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 A-4: 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	-	-

Table A-4: Priors $p(\theta)$.

Parameter	Distribution	Mean	Standard Deviation	Lower Bound	Upper Bound
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).

C.2 Model Equations

$$\mathcal{M}^p = \frac{\epsilon_p}{\epsilon_p - 1}$$

$$\omega = \frac{\epsilon_p}{\Xi \mathcal{M}_t^p}$$

$$\kappa = \omega \left(\sigma + \varphi\right)$$

$$\Psi = \frac{(1-\lambda)\left(1-\delta\left(1-\tau\right)\right)}{\left(1-\lambda+\left(\mathcal{M}_t^p-1\right)\left(1-\lambda\delta\left(1-\tau\right)\right)\right)^2}$$

$$\Phi = \lambda \frac{\left(\sigma+\varphi\right)\Psi}{1-\lambda\gamma}$$

$$\psi_{ya}^n = \frac{1+\varphi}{\sigma+\varphi}$$

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

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

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

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

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

$$y_t^n = \psi_{ya}^n a_t$$

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

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

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

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

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

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

$$y_t = a_t + n_t$$

$$i_t^{ann} = 4\hat{i}_t$$

$$r_t^{r,ann} = 4r_t^r$$

$$r_t^{n,ann} = 4r_t^n$$

$$\pi_t^{ann} = 4\pi_t$$

D Priors for Empirical Application

This section 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 A-5: Additional Priors.

Parameter	Distribution	Mean	Information on standard deviation
μ	Gaussian	0	1 for free elements, 0 otherwise
Σ_u	Inverse Gamma for each element	0.2^2	∞ (scale parameter of 2)
Σ_w	Inverse Gamma for each element	$0.5^2$for GDP, 0.2^2 else	shape parameter is set to half the sample size
$\tilde{X}_0$	Gaussian	0	10
μ^z	Gaussian	0.25 for GDP, 0 else	0.50 for GDP, 0 else