Memoire

Lukas Gahl

1 Introduction

This work is joining 3 strains of literature - Literature on the RBC, the origin of all new keynesian models - Literature on bayesian estimation of structural macroeconomics models - Literature on energy price shocks and their role in the - This refers to the nature of the shock - Shocks as theoretical origin on business cycles are at the origine of this research, identifying suitable shocks that explain more thus adds to the literature on business cycles

The Lucas 1976 critique:

Intertamperol evoluation of macreoconomic variables

Kydland and Prescott (1982) introduced the RBC
- became main model of macroeconomy - rational expectation model

The New Keynesian model extends the RBC by inflation -

Rational expectation models build on two corner stones Chair et al - Structural parameters which are unaffected by policy changes - shocks that have a relevant economic interpretation

Kocherlakota (2007) adjustment principle - the better a model fits to data the better it is for policy advise - in assessing this fit one needs to be wary of over fitting specific data

2 RBC model

The theoretical benchmark model a classic RBC The aim of analysing business cycles

The micro foundation, instead of clear behavioural rules agents optimise their behaviour according to utility optimisation

Conceptual ideas: The efficiency of business cycles (Gali, 2008) in a world of perfect competition and lack of nominal rigidities business cycles might be efficient they are the response to changes in productivity, and are actually the result of a "correcting" force towards an efficient equilibrium this raises questions on the desirability of policy interventions This goes against (Keynes, 1936) who regarded recession as inefficient, due to underutilisation of ressources

Technology shocks Technological shocks for a correctly calibrated model allowed to simulate cycles similar to actual business cycles in first and second order moment (ref). This shed new light on the assumption that technology was solely a driver of long-term growth with neglible impact on business cycles (Gali, 2008, p.3).

RBCs succesfully abstraced from monetary rigidities in their explanation of business cycles

The assumptions of the RBC and their consequences, namely the nonexistance of money and monetary rigidities in the economy are greatly contratsted by emperical evidence (Gali, 2008, p 15)

Assumptions - perfect compitions - flexible prices - technology shocks - infinetly lived households and firms - identical technology across firms

Agents - representative households, where the sum of all households is normed to one - Firms, the representative firm normed to unity

Allocation decisions HH - intertamperol consumption and leasure choice - intertemperol consumption savings

Firm - static optimisation of profit

the model - Non-linearity arises from multiplicative Cobb-Douglas production and additive law of motion of e.g. capital (Campbell) - This raises the need for linear approximations, overview of possible methods: - Models become more complex as researchers are trying to use more realistic functions of utility and incorporate heterogeneity (Taylor Uhlig, 1990) - The method of solving and approximating the model has significant impact on simulated data, hence is relevant when relating models to real data (Taylor Uhlig, 1990)

Methods of linearisation - This work builds on models that can be solved analytically - All models are solved analytically and then log-linearised with 1st order Taylor approximation following (Cambpell) - Based on the log-linearisation the predictions are performed - Log-linear quadratic approximation Rebello (1987) - Under a deterministic singular solution it solves correctly (Cambell) This work follows Campbell and uses the capital stock as the state variable As there is no multiple steady states, this work solves the models anlytically

Criticism the correlation of real wage and hours worked does not correspond to reality, this is the lacmus test of RBC models (Christiano Eichenbaum 1992) "Robert E. Lucas (1981 p. 226) says that "observed real wages are not constant over the cycle, but neither do they exhibit consistent pro- or countercyclical tendencies. This suggests that any attempt to assign systematic real wage movements a central role in an explanation of business cycles is doomed to failure" [CE92]

3 NK

Additional NK assumptions (Gali) - monopolistic competition, inputs are set by private agents according to their own optimisation problem - nominal rigidity, price setting is limited in frequency - this results in short-run neutrality of monetary policy where changes in interest are note directly matched by changes in expected inflation - this is the source of short-run fluctuations - However, in the long-run prices adjust - This causes the response of the economy to be inefficient, unlike RBC where cycles are result of efficient adjustments

NK models are suitable to comparing alternative policy regimes without being subject to the Lucas (1976) critique (Gali)

Small scale new keynesian model solved with first order log-approximation [HS14]

Define stability condition of the linear state-space, Blanchard-Kahn

4 Petrol

High correlation of hours worked and real wage in the RBC model, an fact that is not matched by reality. The aim of RBC research thus must be finding new kinds of shocks, that allow for a more realistic representation of labour supply ()Christiano Eichenbaum 1991).

One such 'new' shock is the the inclusion of petrol as an exogenous shock series [KL92]

Part I
Bringing the model to data

5 The Data

Data sources References for data sources

6 Preprocessing

per capita log HP filter - what is natural GDP - what is cyclicity - why does it yield log-deviations

7 Descriptives

Descriptives Plot data

8 Analysis of models

Compare model covariance matrices, to actual covariance in data

Show theoretical impulse response functions, differences between NK and RBC

9 Estimation methods

As explained in the above the RBC and later NK models are based on structural parameters of the economy. As such they are not subject to the Lucas 1976 critique as their parameters do not directly depend on the choices of

agents (ref). Instead, they represent the deep-rooted dynamics of the dynamic system describing the economy, as such they should exhibit a certain stability over time (ref).

This perspective makes an implicit but important assumption: DSGE model corresponds to the real dynamic system describing the economy. A far reaching assumption which has been considered at least partially validated based on the ability of early DSGE models to generate data matching 1st and 2nd order moments of real variables such as output, consumption and investment (ref). In an attempt to tweak this capability early attempts to parameter to identifying the true parameters of the economy were made. In doing so academics first relied on what [Pre86] referred to as great ratios of the economy. Calculating for example the real interest rate allowed to derive the discount factor β. Other parameters had fewer empirical counterfactuals and where thus not estimated by manually calibrated to sensible values. These inaccuracy gave rise to approaches more deeply rooted in econometric analysis. They were undertaken in the aim of reliably estimating the economy's true structural parameters.

Such early attempts of econometric identification relied on standard frequentists statistics such as the maximum-likelihood estimators (MLE) (ref). Another approach pioneered by Blanchard & Kahn (ref) referred to as simulation method of moments, later progressed into the general method of moments (GMM) attempted to minimize the distance between simulated and real data moments. The GMM allowed to discriminate between several competing specification of the same structural model [CE92]. Comparing the estimated parameters of the model with real data values then served as inspection of model fit to data. However, as models grew in size a major short-coming of GMM became apparent. The method of moment conditions were no longer sufficient to estimate model fit across the variety of parameters [GQN13]. Likewise, the ML estimator turned out to be limited in its ability to estimated the entirety of parameters, requiring manual calibration of some parameters [GQN13].

While frequentists statistics knew to overcome the identification problem in introducing theoretical moments Smith (1993) it was soon challenged by a different perspective. At the heart of all frequentist statistics lies the assumption that a given model corresponds to the true structural process, it is the 'true' model (ref). If this assumption cannot be met all estimation will be biased and thus need to be corrected for. Again remedies where found (ref) Smith (1993) or Dridi, Guay, and Renault (2007) in varying parameters most relevant to overall model moments and keeping less relevant ones constant throughout econometric estimation. Yet it is argued that this approach at least partly failed in overcoming the conditional bias introduced

by non-relevant parameters [GQN13].

Consequently, a new philosophy of linking models to data was identified. Bayesian statistics replaces the assumption of a single true model by the convenient formulation of conditional likelihood making it more appealing to the purposes of DSGE modelling [GQN13]. In doing so the Bayesian statistics relies on Bayes law linking the conditional likelihood of an event to the prior distribution.

$$P(\Theta|Y_T) = L(Y_T|\Theta)P(\Theta)$$

A given model specification can be linked to a likelihood evaluated against the backdrop of data the system is meant to generate (ref). This allows to assess a model's suitability. The frequentist's idea of a single true parameter is replaced by distributions assigning different probabilities to parameter values. These prior distribution reflect previous knowledge allowing to narrow down the possible space of parameters. Instead of identifying a true parameter Bayesian statistics, aims at evaluating these believes and at decreasing uncertainty around the parameter space i.e. the variance of the distribution. This process translates the prior distribution into a posterior distribution, not necessarily belonging to the same class. This process of evaluation is referred to as sampling and builds on numerical and stochastic calculus to approximate the true posterior distribution (ref). Furthermore, Bayesian statistics naturally provides great forecasting properties. Once uncertainty around the space of the real system has been reduced it allows for make informed forecasts, naturally providing confidence bans.

The following section will therefore be concerned with these four concepts, conditional likelihood, prior and posterior distribution, sampling and forecasting. DSGE literature has developed a rich body of literature covering the case of the canonical DSGE model rather well (ref). The state of research is currently concerned with the estimation of larger models and non-linear and not normally distributed posterior distributions (ref). However, given the simplicity of the above discussed models current techniques are more than sufficient for this thesis purposes.

9.0.1 The log-likelihood

Bayesian analysis of DSGE models is concerned with the transition from a prior distribution reflecting previous beliefs to the true posterior distribution (ref). This process relies on Bayes law as illustrated above, which requires the prior distribution and the marginal likelihood of data given the representation of the system $L(Y|\Theta)$ (ref). This section is concerned with obtaining the latter.

The marginal likelihood for a linear state system requires a technique allowing to assess the fit of data to the system. In doing so one relies on the fact that DSGE models when expressed as linear dynamic systems are data generating processes. This allows to formulate the question of how likely a given set of data has been generated by the system. A method allowing to do so is the Kalman filter (ref). This filter originates from engineering and has originally been developed to track the position of an air plane off signals from different sensor. As sensor signals are inherently noisy the filter is designed to exploit one's knowledge about the measurement noise in order to reduce the variance on its believe about the actual position or state of the plane. In doing so the filter imposes the important assumption of Gaussian distributed noise, exploiting the fact that any multiple of two Gaussian distributions will again yield a Gaussian distribution. The resulting Gaussian then has a lower variance than its parents, which is equivalent to saying that the resulting distribution has narrowed down the believe on the current state of the system. This assumption is especially important for the purposes of DSGE modelling, as the likelihood function of the Kalman filter relies on the assumption of Gaussian noise. Applied to DSGE models this implies that the joint distribution of shocks must be Gaussian [HS14]. This is usually satisfied for small scale New Keynesian models if the individual shocks are Gaussian (ref). For larger DSGE models this assumption does not necessarily hold and shocks are highly correlated. As mentioned above the is state of research and demands other methods of evaluation such as the Kalman-Particle filter (ref). As this work is concerned with the canonical model the Kalman filter will be sufficient.

Given such linear Gaussian system the Kalman filter becomes especially desirable as it provides optimal forecast (ref), making any other method of evaluation obsolete. The below will now briefly explain how the Kalman filter is constructed off the linear system and end with deriving its likelihood function.

Given the generic linear system below wit $\mathbf{F}$ as the transition and $\mathbf{R}$ as the shock noise matrix the Kalman filter proceeds in two steps, a predict and an update step:

$$\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{R}\epsilon_t$$

$$\epsilon_t \sim N(0, \sigma_\epsilon)$$

Predict step

In the predict step the Kalman filter propagates the system of state variables $\mathbf{X}_t$ according to the transition matrix $\mathbf{F}$. This matrix defines

the system and as such is time independent. The propagation then yields an estimate for the X_{t+1} system state. This could be supplemented by directions or external influences on the system that one is aware of. However, this is not the case for this work.

$$\hat{\mathbf{X}}_{k|k-1} = \mathbf{F}\mathbf{X}_{k-1|k-1}$$

The second key aspect to the Kalman Filter is the system state variable's covariance matrix $\mathbf{P}_t$, which is also time dependent. As pointed out in the above, this matrix is assumed to be Guassian, thus semi-definit positive (ref). The covariance matrix and the state $\mathbf{X}$ essentially build a Guassian distribution summarising the uncertainty around the actual state.

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q}$$

Filter step

Once the predicted state $\mathbf{X}$ and covariance matrix $\mathbf{P}$ have been determined the Kalman filter proceeds to identifying the most likely system state. For this it takes into account the state variables measurement referred to as $\mathbf{z}_t$, which due to measurement noise is not considered to always reflect the actual actual state of the system. The actual state is instead thought of to ly in-between the predicted state $\mathbf{X}_{k|k-1}$ and the measurement $\mathbf{z}_k$. The objective of the filtering step is therefore to identify this true state in-between the two. For this purpose the Kalman filter derives the Kalman gain $\mathbf{K}_{k|k-1}$, which can be though of as a weighting function between $\mathbf{X}_{k|k-1}$ and $\mathbf{z}_k$. In building this average the filter takes into account the process noise and the covariance matrix as well as the accuracy of past iterations.

$\mathbf{y}_k$ is the difference between measurement and prediction. The matrix $\mathbf{H}$ is the measurement matrix, ensuring that predicted state $\mathbf{X}_{k|k-1}$ and $\mathbf{z}_k$ are in the same units.

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{H}\mathbf{X}_{k|k-1}$$

The $\mathbf{S}_k$ captures the filters uncertainty by combining the covariance around the predicted state $\mathbf{X}_{k|k-1}$ with the process noise $\mathbf{R}$. In DSGE terminology the process noise is derived from the shock covariance matrix (ref).

$$\mathbf{S}_k = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}$$

The uncertainty term $\mathbf{S}_k$ is then used to calculate the weighting function between measurement and predicted state the Kalman gain.

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_k^{-1}$$

The filtered state which will be next iterations start then is calculated off the Kalman gain and $\mathbf{y}_k$. The covariance matrix is updated in a similar fashion.

This yields a new narrower distribution summarising the uncertainty around the true systems state.

$$\mathbf{X}_{k|k} = \mathbf{X}_{k|k-1} + \mathbf{K}_k \mathbf{y}_k$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H})\mathbf{P}_{k|k-1}$$

Likelihood function

Having introduced the Kalman filter algorithm the setting of the starting values remain. For the Kalman filter these are less decisive, has the filter is very capable in converging to the true state. However, as the models at hand are describing deviations from the steady state this work has set $\mathbf{X}_{0|0}$ to zero, assuming that the system is in its steady state to begin with, an approach generally followed by the literature. Secondly, this section has initially been dedicated to the Kalman filter's likelihood function. Having explored the propagation one can now derive the likelihood function. This function is heavily reliant on the assumption of Gaussian system noise, as it draws from the normal distribution in evaluating the likelihood for each iterations. In the below when referring to likelihood this thesis will however be referring not to the individual likelihood of each iteration but to the cumulated likelihood across all iterations.

The probabiltiy desinty function for any normal distribution is given by:

$$\mathcal{L} = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}\frac{x-\mu}{\sigma}]^2$$

The Kalman filter relies on the multivariate where the standard deviation is given by system uncertainty $\mathbf{S}_k$ and the difference between mean and measurement are expressed through $\mathbf{y}_k$.

$$\mathcal{L}_k = \frac{1}{\mathbf{S}_k\sqrt{2\pi}} \exp[-\frac{1}{2}\mathbf{y}_k^T \mathbf{S}_k^{-1} \mathbf{y}_k]$$

This work has drawn from filterpy in order to perform the above calculations (ref).

9.0.2 Sampler

There are local and global identification problems

The key aspect of Bayesian estimation across DSGE models is the identification of the true parameter distribution, the posterior. However, mapping a DSGE model's parameters to its posterior is non-linear in the parameter vector Θ. Therefore the posterior distribution cannot be evaluated analytically [HS14]. Instead a numerical approximation mechanism is require, this is referred to as the sampler [GQN13]. The Metropolis-Hastings Monte

Carlo Markov Chain (MH-MCMC) sampler is the predominant sampling method in linear Bayesian estimation literature [GQN13]. The MH-MCMC sampler is applicable to small scale New Keynesian models, for which it usually delivers good results as the joint parameter distribution is well-behaved and elliptical [HS14]. This reduces the posterior identifiaction problem to a global problem. Larger NK models exhibit no-eliptic parameter distributions and thus requirer samplers which are able to distinguish local and global likelihood maxima in their identification of the posterior [HS14]. As this thesis is concerned with the small scale canoncial model the Metropolis-Hastings sampler is sufficiently accurate and will be used.

The MH-MCMC sampler suggests new posterior candidates according to a multivariate random walk. The posterior is extend by a suggestion if it provides an improvement in the likelihood as described in (section). The algorithm generates a stable Markov chain, which is meant to exhibit low autocorrelation and low variance of the MH estimator. If this is fullfilled the resulting chain is equivalent to the posterior distribution [HS14].

Metropolis Hastings sampling algorithm - it is important to note that a draw from the shocks prior, factors into the drawing process - procedure also followed by, with 300,000 draws of which 20,000 are considered burn in [CL19]

The below will be explaning this thesis implementation of the MH sampler and methods of evaluating its convergence to the true posterior distribution. The algorithm proceeds in three main steps. First a candidate for the posterior $\hat{\mathbf{\Theta}}_{k|k-1}$ is suggested based on the random-walk law of motion. The random-walk departs from the last accepted posterior candidate and suggests a new candidate by departing from the $\mathbf{\Theta}_{k-1|k-1}$ according to the zero-mean Gaussian distribution ϵ. The distribution's variance $\mathbf{\Sigma}$ can be set to either the identity matrix $\mathbf{I}$ or to contain the parameter prior variances on its diagonal. The latter is recommended if well-defined priors are at play [HS14]. To complete the random-walk the MH sampler also employs a gain parameter η that essentially scales the distance of the new posterior candidate away from the last accepted posterior $\mathbf{\Theta}_{k-1|k-1}$. This value has been calibrated by several studies, which recommend an $\eta = 0.234$ for the special case of a multivariate normal distribution (ref). Generally the value of η should correspond to an acceptance rate between 20% and 40% of suggested draws [HS14].

$$\hat{\mathbf{\Theta}}_{k|k-1} = \mathbf{\Theta}_{k-1|k-1} + \eta\epsilon$$

$$\epsilon = N(0, \mathbf{\Sigma})$$

The posterior candidate $\hat{\mathbf{\Theta}}_{k|k-1}$ is th evaluated base on the Kalman Filter likelihood function introduced in (section). This yields the likelihood of

posterior candidate given the data $\mathbf{Y}$.

$$\mathcal{L}(Y|\hat{\mathbf{\Theta}}_{k|k-1})$$

Once the likelihood is optained the alogrihtm proceeds into the two step acceptance procedure. First the likelihood of the candidate is compared to last accepted posterior in order to determine whether this draw constitutes an improvement. This is done according to Bayes' law.

$$\omega_k = \min\{\frac{\mathcal{L}(Y|\hat{\mathbf{\Theta}}_{k|k-1}) * P(\hat{\mathbf{\Theta}}_{k|k-1})}{\mathcal{L}(Y|\mathbf{\Theta}_{k-1|k-1})P(\mathbf{\Theta}_{k-1|k-1})}, 1\}$$

The algorithm proceeds as follows [HS14]

Posterior Θ, prior p, suggestions θ

1. A new draw ζ is suggested based on a random draw from the prior distribution of parameters Θ. In order to determine whether this draw is admissible to the posterior $\Gamma_{1:i}$ the following steps are followed.

   $$\zeta = \Theta_{i-1} + \eta$$

   $$\eta \sim N(0, c\hat{\Sigma})$$

   [HS14] recommends to set Σ either to the identity matrix or as a diagonal matrix with the posterior variances on its diagonal.

   The scaling factor c is under the assumption of Gaussian normal joint posterior distribution usually put between .2 and .4

2. The new draw ζ is evaluated in likelihood using the DSGE model's likelihood function given the data Y, yielding $p(Y|\zeta)$.

3. In a two stage procedure the algorithm then either accepts or rejects the new draw ζ as part of the Markov chain of posterior distributions.

4. First the posterior likelihood of ζ is calculated based on Bayes theorem. It is then compared to the the posterior likelihood of the last accepted draw, that is now part of the posterior Γ_{i-1}.

5. The second step of the decision algorithm draws from a uniform random variable $\psi \sim U(0, 1)$. The new draw ζ is accepted if its likelihood ration ω_i is accepted if $\omega_i \geq \phi$. This step is equivalent to moving to a higher point on the likelihood surface [HS14]

6. If accepted ζ will become Γ_i

7. The algorithm is then repeated

How many runs

As illustrated in the above MH-MCMC is a recursive method and as such requires an initial value for the posterior Γ_0. This values is usually obtained throughout a so-called burn in period corresponding to 10 % of total iterations (ref). The total number of iterations depends on the

Moreover, once the algorithm has run one needs to evaluate this performance. A method beyond visual inspection is suggested by (ref).

The MH random walk is highly suceptible to mis specification of the posterior of the prior distribution is far away from the true posterior [HS14]

9.1 Priors

The assumption for prior selection is that priors are obtained from some past knowledge. If this knowledge is independent of the data used to evluate the liklihood function than the prior holds. This is fullfiled if the data for prior building at least predates the data for likelihood evluation [HS14].

Priors are statistical distributions which either reflect believes about the strucutral mechanism or are derived from data, external to the estimation process [DNS08] - Priors thus refelect microeconomic evidence, for example regarding labour supply and elasticity - However, they are a simplication of the real joint likelihood function of the strucutural model. Such function is difficult and sometimes impossible to evaluate

Priors [DNS08] - Priors are usually assumed to be independent for simplicities sake - Priors are often calibrated for the model to match the covariance of real data

[DNS08] suggest two kinds of data sources. Pre-sample data, which is not included in the model evaluation process or in-sample data, that is explicitly excluded from the model evaluating likelihood function. As the model parameters are thought of as structural, hence relatively time-invariant, more dated data can be used to estimate them.

Priors are divided into three groups, following [DNS08]

9.1.1 Steady state priors

The 'great ratios' or long-run relationships as [KP82] referred to them pin down the steady state

$$\theta_{ss} = \{\alpha, \beta, \delta, \pi^*, \eta_p\}$$

When determining the steady-state priors θ_{ss} and assuming independence it is possible that the resulting joint prior distributions assigns high probability to unrealistic steady state values [DNS08]. To circumvent this issue other method

This priors are usually estimated from pre-sample data [HS14] this work follows this approach and thus

9.1.2 Exogenous priors

The exogenenous propagation paraemeters these are the exogenenous shock parameters, their autoregressive coefficient and their standard deviation

$$\theta_{exog} = \{\rho_a, \rho_s, \sigma_a, \sigma_s\}$$

Exogenous priors can be evaluated from the model specification in keeping the influence of other priors minimal. Del Negro suggests, to set all other parameters to zero where possible. The system of equations then reduces to variables that are depend on the exogenous shocks. This reduced system can then be used to evaluate the likelihood of prior values, using out-of sample data of the later calibration. An evaluation based on the likelihood of the reduced system is not necessary if the variation in observable variables can be directly retraced to individual shocks. In a simple model where shocks do not intersect in their influence on observable variables, e.g. in the case of a singular technology shock the variance of the shock can be directly inferred from the observable variable. In case of intersecting multiple shocks the likelihood of the reduced system needs to be consulted to choose appropriate priors, as their influence on observables cannot be directly inferred.

In case of exogenous observable shocks, such as energy prices, the above procedure is not required. Other than the level of technology energy price shocks are measurable, wherefore one can rely on the observable first and second order moments (ref).

9.1.3 Endogenous priors

The endogenous propagation parameters, including the shock autoregressive coefficients are parameters involved with the propagation of shocks through the system. These parameters heavily rely on micro-evidence from external data sets [DNS08] All remaining parameters are stacked into the the following

$$\theta_{endog} = \{\sigma_C, \sigma_L, \}$$

} Priors of the remaining variables are chosen in order to reflect believes or micro evidence. This work draws from the choice of [DNS08] in their endogenous priors, given the similarity in models

Priors are chosen in order to match first and second order moments individually as well as in a joint distribution across all parameters [DNS08]

9.1.4 Reference Model BVar

[Sch00]

Table 1: Prior distributions

model param	mean NK	NK Energy	RBC	std NK	NK Energy	RBC
alpha	0.285714		0.285714	0.02551		0.02551
alpha_m		0.285714			0.02551	
alpha_n		0.285714			0.02551	
epsilon		0.5			0.073529	
epsilon_A	0.099099		0.099099	0.007378		0.007378
epsilon_R	0.099099			0.007378		
epsilon_T	0.099099			0.007378		
epsilon_Y	0.099099			0.007378		
epsilon_pi	0.099099			0.007378		
epsilon_s		0.099099			0.007378	
eta_p	0.746269			0.013149		
gamma_R	2.0			1.0		
gamma_Y	2.0			1.0		
gamma_pi	2.0	2.0		1.0	1.0	
mu		9.440416			56.352936	
sigma_C	2.0	2.0	2.0	4.0	4.0	4.0
sigma_L	2.0	2.0	2.0	4.0	4.0	4.0

BVar with normally indpendently distributed lag parameters [CL19]

This work uses a frequentist VAR to determine the optimal number of lags, which are found to be 4. This corresponds to the choice of four lags in [CL19]

9.2 Forecasts

Relevant to policy recommendations How well can models capture the dynamics of the economy, e.g. as a spectrum of possible outcomes

Including priors into bayesian VAR makes the model better at forecasting [CL19]

To compare models this work follows a bayesian averaging approach as in [CL19] - (Stock Watson 2003) argue that a simple average of models provides the best forecast - ()Jacobson and Karlsson (2004) and Wright (2009)) argue in favour of BMA

BMA - the weights of each model are calculated from its posterior model probability [CL19] - This is traditionally done for structurally similar models in order to discriminate between different priors - the posterior probability is thereby obtained from [CL19]

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_K^{k=1} py(y|M_k)p(M_k)}$$

$$p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k$$

where $p(M_k)$ is the prior density, and $p(y|M_k)$ is the model's marginal likelihood

Anderson (2008) suggest to use the predictive likelihood instead of the marginal model likelihood

"Simply speaking, our Bayesian combination forecast (point forecast) is a weighted average of the K individual point forecasts, wherein the weights are determined by their predictive likelihoods" [CL19]

Comparing forecasting performance Diebold-Mariano test for point forecast comparison following [CL19] - MSE based point forecast comparison Amisano–Giacomini test for density forecast performance following [CL19]

References

[CE92] Lawrence J. Christiano and Martin Eichenbaum. Current real-business-cycle theories and aggregate labor-market fluctuations. *The American Economic Review*, pages 430–450, 1992. Publisher: JSTOR.

[CL19] Kuo-Hsuan Chin and Xue Li. Bayesian forecast combination in VAR-DSGE models. *Journal of Macroeconomics*, 59:278–298, March 2019.

[DNS08] Marco Del Negro and Frank Schorfheide. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics*, 55(7):1191–1208, October 2008.

[GQN13] Pablo A. Guerrón-Quintana and James M. Nason. Bayesian estimation of DSGE models. In *Handbook of research methods and applications in empirical macroeconomics*, pages 486–512. Edward Elgar Publishing, 2013.

[HS14] Edward Herbst and Frank Schorfheide. Bayesian inference for DSGE models. *Manuscript. Board of Governors of the Federal Reserve System, Washington, DC*, 2014.

[KL92] In-Moo Kim and Prakash Loungani. The role of energy in real business cycle models. *Journal of Monetary Economics*, 29(2):173–189, April 1992.

[KP82] Finn E. Kydland and Edward C. Prescott. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370, 1982. Publisher: JSTOR.

[Pre86] Edward C. Prescott. Theory ahead of business-cycle measurement. In *Carnegie-Rochester conference series on public policy*, volume 25, pages 11–44. Elsevier, 1986.

[Sch00] Frank Schorfheide. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics*, 15(6):645–670, 2000. Publisher: Wiley Online Library.

Part II
Appendix

10 Appendix A

10.1 RBC