

Memoire

Lukas Gahl

May 18, 2023

Abstract

Contents

1	Introduction	4
2	Literature Review	5
3	DSGE Models	6
3.1	RBC model	6
3.2	NK	7
3.2.1	Consumption bundle	7
3.2.2	Household	8
3.2.3	The firm	8
3.2.4	Price setting	8
3.3	Petrol	9
4	Data	10
4.1	Preprocessing	10
4.2	Analysis of models	11
5	Bayesian Estimation	12
5.1	The log-likelihood	14
5.2	Sampler	17
5.3	Priors	19
5.3.1	Steady state priors	20
5.3.2	Exogenous priors	20
5.3.3	Endogenous priors	21
6	Forecast evaluation	22
6.1	Reference Model BVar	22
6.2	Model Forecasts	22
7	Conclusion	24

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 evolution 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 Literature Review

3 DSGE Models

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 (Campbell) - Based on the log-linearisation the predictions are performed - Log-linear quadratic approximation Rebello (1987) - Under a deterministic singular solution it solves correctly (Campbell) 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 analytically

3.1 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 under-utilisation of resources

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 negligible impact on business cycles (Gali, 2008, p.3).

RBCs successfully abstracted from monetary rigidities in their explanation of business cycles

The assumptions of the RBC and their consequences, namely the non-existence of money and monetary rigidities in the economy are greatly contrasted by empirical evidence (Gali, 2008, p 15)

Assumptions - perfect compitions - flexible prices - technology shocks - infinitely 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 - intertemporal consumption and leisure choice - intertemporal 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)

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" [Christiano and Eichenbaum, 1992]

3.2 NK

3.2.1 Consumption bundle

$$\max_{C_{it}} \left(\int_0^1 C_{it}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}}$$

s.t.

$$\int_0^1 P_{it} C_{it} di \leq Z_t$$

solving this results in an expression for the price

$$P_t = \left(\int_0^1 P_{it}^{1-\epsilon} di \right)^{\frac{1}{1-\epsilon}} \quad (1)$$

with further manipulation it can be shown that

$$C_{it} = \left(\frac{P_{it}}{P_t} \right)^{-\epsilon} C_t$$

the share of a good in the consumption bundle is determined by its relative price and the elasticity of substitution

3.2.2 Household

Having found C_t composition the household now needs to decide the share of income that it will allocate to consumption

$$\max_{B_t, C_t, L_t} \sum_{t=0}^{\infty} \beta^t U(C_t, L_t)$$

s.t.

$$P_t C_t + Q_t B_t = B_{t-1} + W_t N_t + T_t$$

FOC:

$$-\frac{U_{n,t}}{U_{c,t}} = \frac{W_t}{P_t} = C_t^\sigma N_t^\phi$$

$$Q_t = \beta \mathbb{E}_t \left[\frac{C_{t+1}^\sigma}{C_t} \frac{P_t}{P_{t+1}} \right]$$

3.2.3 The firm

$$\max_{N_t} = A_t N_t^{1-\alpha}$$

s.t.

$$P_t Y_t = W_t N_t$$

FOC

$$W_t = (1 - \alpha) A_t N_t^{-\alpha} P_t$$

3.2.4 Price setting

In price setting we return to the definition of the price as shown in (section) Firms individually decide their prices However, they are only able to reset them at rate θ . As the number of firms is normed to unity θ serves as the share of firms that reset their price in a given period t . This allows to derive an expression for the price level t

$$P_t = \left[\theta P_{it-1}^{1-\epsilon} + (1 - \theta) \hat{P}_t^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}}$$

$$\Pi_t^{1-\epsilon} = \theta + (1 - \theta) * \left(\frac{\hat{P}_t}{P_{t-1}} \right)^{1-\epsilon} \quad (2)$$

Having derived an expression for inflation this section will now turn to the heart of the NK model, the optimal price setting by firms.

$$\max_{\hat{P}_t} \sum_{k=0}^{\infty} \theta^k \mathbb{E}_t \left[Q_{t,t+k} \left(\hat{P}_t Y_{t+k|t} - TC_{t+k|t}^n(Y_{t+k|t}) \right) \right]$$

s.t.

$$Y_{t+k|t} = \left(\frac{\hat{P}_t}{P_{t+k}} \right)^{-\epsilon} C_{t+k}$$

where $Q_{t,t+k}$ is the stochastic discount factor FOC:

$$\sum_{k=0}^{\infty} \theta^k \mathbb{E}_t \left[Q_{t,t+k} Y_{t+k|t} \left((1 - \epsilon) + \epsilon \frac{MC_{t+k|t}^n}{\hat{P}_t} \right) \right]$$

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 not 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 [Herbst and Schorfheide, 2016]

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

3.3 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 inclusion of petrol as an exogenous shock series [Kim and Loungani, 1992]

4 Data

All data is gathered from St. Louis Federal Reserve data base (ref), the specific data code can be found in the (appendix). As such all data has been seasonally adjusted already and as pointed out by [Pfeifer, 2021] is if revised in hindsight adjusted in its entirety so that all measurements depart from the same assumption.

The majority of the data used in across this thesis is standard to the DSGE modelling literature. However, some adjustments have been made to account for the inclusion of a petrol sector. With regards to inflation π this work relies on the median consumer price index for the RBC and the NK model. However, as the aim of the NK energy model is to analyse inflation using the CPI would be counterproductive. The model itself should generate an inflation that corresponds to the CPI as it includes the petrol sector into its dynamics. In order to accurately reflect on the models ability to do so one therefore has to remove energy price inflation from the data. Only doing so allows to assess the models real ability to replicate inflation. The NK energy model will therefore rely on the "Sticky Price Consumer Price Index less Food and Energy". A similar logic applies to consumption expenditure which for the NK energy model does not include expenditure for energy and utilities.

The other variables used for the evaluation of model suitability are rather standard. The nominal interest rate is proxied by the federal funds effective rate. Consumption and investment are measured by household expenditure data. Labour and wage refer to non-farm business total hours worked as well as average hourly wage as recommended for the US economy (ref). As proxy of energy cost this thesis employs the WTI Crude Oil Index as suggested by (ref) as it is available for a longer time horizon as opposed to the more recent Brent Crude Oil Index.

4.1 Preprocessing

Economic data is usually considered in per capita terms, in order to account for different population dynamics (ref). The same procedure will be applied across this work. However, population dynamics are not constantly collected wherefore its records display significant irregularities and not the smooth dynamic of true population growth or decline [Pfeifer, 2021]. Using these error prone records thus will distort the data signal

of other more carefully collected economic data. In order to circumvent this irregularity a smoothing with the Hodrick-Prescott (HP) filter's (at level 10,000) is suggested by (ref) Edge, Gürkaynak, and Kisacikoglu (2013). This work follows this approach and divides trending variables by population data to obtain per capita variables.

However, population dynamics are only one source of trend in variables. Variables measured in monetary terms are also concerned by inflation. [Pfeifer, 2021] (Grohe, Uribe) recommend de-trending nominal variables by the GDP deflator in order to account for this. This work follows this approach and relies on inflation corrected data so that monetary variables are expressed in the same quantities.

As the model studied across this thesis are describing log-deviations of economic variables from their respective steady states further transformation is required. This traditionally involves separating the underlying long-trend of GDP per capita from its cycle variations. A commonly employed method for doing so is the Hodrick-Prescott filter, which if fed with log-transformed variables will output the desired log-deviations from the trend as well as the trend itself (ref). As this work is concerned with quarterly variables the smoothing parameter has been set to 1600 as suggested by the work of [Ravn and Uhlig, 2002]. This procedure by construction yields stationary time series. In order to confirm the stationarity this work has again as commonly the case confirmed stationarity of all variables using the Augmented-Dickey Fuller test (ref).

To this point all trending variables have been transformed into stationary time series of log deviation from their respective natural level. The percentage variables, namely inflation and interest rate require no such treatment. However, as they are stated in annual terms they require translation into quarterly variables. This has been done according to the following formula $(1 + \frac{\mathbf{X}}{100})^{1/4} - 1$.

An overview over the transformed variables can be found in the (appendix). The descriptives statistics can be found in the (appendix).

4.2 Analysis of models

Compare model covariance matrices, to actual covariance in data Show theoretical impulse response functions, differences between NK and RBC

5 Bayesian Estimation

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 [Prescott, 1986] 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 were 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 [Christiano and Eichenbaum, 1992]. 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 [Guerrón-Quintana and Nason, 2013]. Likewise, the ML estimator turned out to be limited in its ability to estimated the entirety of parameters, requiring manual calibration of some parameters [Guerrón-Quintana and Nason, 2013].

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 were 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 [Guerrón-Quintana and Nason, 2013].

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 [Guerrón-Quintana and Nason, 2013]. 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 beliefs 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 bands.

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.

5.1 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 [Herbst and Schorfheide, 2016]. 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:

$$\begin{aligned}\mathbf{X}_t &= \mathbf{F}\mathbf{X}_{t-1} + \mathbf{R}\epsilon_t \\ \epsilon_t &\sim N(0, \sigma_\epsilon) \\ \begin{bmatrix} y_t \\ z_t \end{bmatrix} &= \begin{bmatrix} \rho_y & \rho_{yz} \\ \rho_{zy} & \rho_z \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \\ \begin{bmatrix} \epsilon_{y,t} \\ \epsilon_{z,t} \end{bmatrix} &\sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_z^2 \end{bmatrix} \right)\end{aligned}$$

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 state of the system. The actual state is instead thought of to lie 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 thought 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 filter's 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 system's 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 probabilitiy desinty function for any normal distribution is given by:

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

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 .

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

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

5.2 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 [Herbst and Schorfheide, 2016]. Instead a numerical approximation mechanism is require, this is referred to as the sampler [Guerrón-Quintana and Nason, 2013]. The Metropolis-Hastings Monte Carlo Markov Chain (MH-MCMC) sampler is the predominant sampling method in linear Bayesian estimation literature [Guerrón-Quintana and Nason, 2013]. 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

[Herbst and Schorfheide, 2016]. This reduces the posterior identification problem to a global problem. Larger NK models exhibit no-elliptic parameter distributions and thus require samplers which are able to distinguish local and global likelihood maxima in their identification of the posterior [Herbst and Schorfheide, 2016]. As this thesis is concerned with the small scale canonical 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 extended 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 fulfilled the resulting chain is equivalent to the posterior distribution [Herbst and Schorfheide, 2016].

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 [Chin and Li, 2019]

The below will be explaining 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 $\hat{\Theta}_{k|k-1}$ 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 $\Theta_{k-1|k-1}$ according to the zero-mean Gaussian distribution ϵ . The distribution's variance Σ 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 [Herbst and Schorfheide, 2016]. 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 $\Theta_{k-1|k-1}$. 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 [Herbst and Schorfheide, 2016].

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

$$\epsilon = N(0, \Sigma)$$

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

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

Once the likelihood is obtained the algorithm 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.

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

The likelihood ratio is then fed into the second step the random acceptance. The above obtained ω_k is compared the uniform variable $\phi \sim U(0, 1)$. If $\omega_k \leq \phi_k$ the draw is accepted and the candidate is assumed into the posterior. This process is repeated N times and thus generates the posterior distribution. As illustrated in the above MH-MCMC is a recursive method and as such requires an initial value for the posterior Θ . This value 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 speed of convergence of the chain but usually ranges between 20,000 and 50,000 iterations (ref).

The following will now describe the evaluation steps to insure convergence of the chain. 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 susceptible to mis specification of the posterior of the prior distribution is far away from the true posterior [Herbst and Schorfheide, 2016]

5.3 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 evaluate the likelihood function than the prior holds. This is fulfilled if the data for prior building at least predates the data for likelihood evaluation [Herbst and Schorfheide, 2016].

Priors are statistical distributions which either reflect beliefs about the structural mechanism or are derived from data, external to the estimation process [Del Negro and Schorfheide, 2008]

- Priors thus reflect microeconomic evidence, for example regarding labour supply and elasticity - However, they are a simplification of the real joint likelihood function of the structural model. Such function is difficult and sometimes impossible to evaluate

Priors [Del Negro and Schorfheide, 2008] - Priors are usually assumed to be independent for simplicities sake - this is true for shocks in small scale canonical model, as shocks are assumed to be independent, an important assumption for the bayesian analysis to hold [Herbst and Schorfheide, 2016] - Priors are often calibrated for the model to match the covariance of real data

[Del Negro and Schorfheide, 2008] 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 [Del Negro and Schorfheide, 2008]

5.3.1 Steady state priors

The 'great ratios' or long-run relationships as [Kydland and Prescott, 1982] 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 [Del Negro and Schorfheide, 2008]. To circumvent this issue other method

This priors are usually estimated from pre-sample data [Herbst and Schorfheide, 2016] this work follows this approach and thus

5.3.2 Exogenous priors

The exogenous propagation parameters these are the exogenous 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).

5.3.3 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 [Del Negro and Schorfheide, 2008] 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 [Del Negro and Schorfheide, 2008] 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 [Del Negro and Schorfheide, 2008]

6 Forecast evaluation

6.1 Reference Model BVar

[Schorfheide, 2000]

BVar with normally independently distributed lag parameters [Chin and Li, 2019]

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 [Chin and Li, 2019]

6.2 Model 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 [Chin and Li, 2019]

To compare models this work follows a bayesian averaging approach as in [Chin and Li, 2019]

- (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 [Chin and Li, 2019] - This is traditionally done for structurally similar models in order to discriminate between different priors - the posterior probability is thereby obtained from [Chin and Li, 2019]

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{K=1}^K p(y|M_k)p(M_k)}$$

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

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" [Chin and Li, 2019]

Comparing forecasting performance Diebold-Mariano test for point forecast comparison
following [Chin and Li, 2019] - MSE based point forecast comparison Amisano–Giacomini
test for density forecast performance following [Chin and Li, 2019]

7 Conclusion

Bibliography

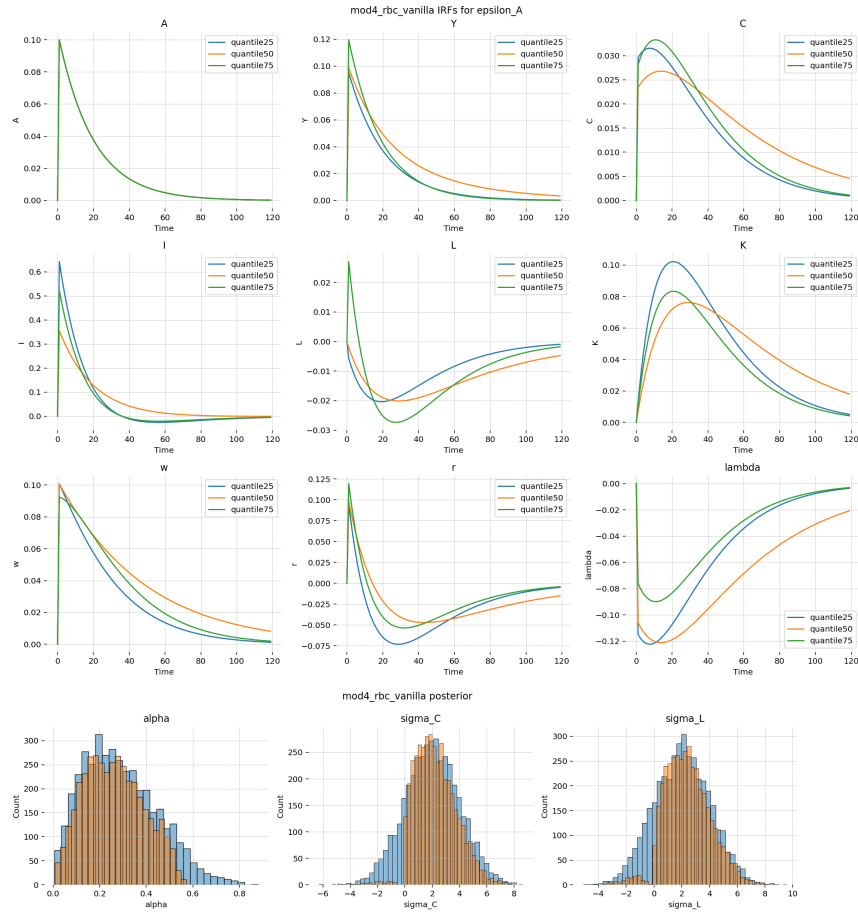
References

- [Chin and Li, 2019] Chin, K.-H. and Li, X. (2019). Bayesian forecast combination in VAR-DSGE models. *Journal of Macroeconomics*, 59:278–298.
- [Christiano and Eichenbaum, 1992] Christiano, L. J. and Eichenbaum, M. (1992). Current real-business-cycle theories and aggregate labor-market fluctuations. *The American Economic Review*, pages 430–450. Publisher: JSTOR.
- [Del Negro and Schorfheide, 2008] 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.
- [Guerrón-Quintana and Nason, 2013] Guerrón-Quintana, P. A. and Nason, J. M. (2013). Bayesian estimation of DSGE models. In *Handbook of research methods and applications in empirical macroeconomics*, pages 486–512. Edward Elgar Publishing.
- [Herbst and Schorfheide, 2016] Herbst, E. P. and Schorfheide, F. (2016). *Bayesian estimation of DSGE models*. Princeton University Press.
- [Kim and Loungani, 1992] Kim, I.-M. and Loungani, P. (1992). The role of energy in real business cycle models. *Journal of Monetary Economics*, 29(2):173–189.
- [Kydland and Prescott, 1982] Kydland, F. E. and Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370. Publisher: JSTOR.
- [Pfeifer, 2021] Pfeifer, J. (2021). A Guide to Specifying Observation Equations for the Estimation of DSGE Models.
- [Prescott, 1986] Prescott, E. C. (1986). Theory ahead of business-cycle measurement. In *Carnegie-Rochester conference series on public policy*, volume 25, pages 11–44. Elsevier.
- [Ravn and Uhlig, 2002] Ravn, M. O. and Uhlig, H. (2002). On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *Review of Economics and Statistics*, 84(2):371–376.

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

Appendix

Appendix A





Appendix B

Table 1: Prior distributions

model param	mean			std		
	NK	NK Energy	RBC	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_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

Table 2: St. Louis Fed Data Variables

	Variable	Data variable FRED Code	Data variable description
0	Y	GDP	Gross Domestic Product
1	Y_p	GDPPOT	Real Potential Gross Domestic Product
2	pi_s	CORESTICKM159SFRBATL	Sticky Price Consumer Price Index less Food an...
3	pi_c	MEDCPIM158SFRBCLE	Median Consumer Price Index (core inflation)
4	r	FEDFUNDS	Federal Funds Effective Rate
5	I	GPDI	Gross Private Domestic Investment
6	C	PCEPILFE	Personal Consumption Expenditures Excluding Fo...
7	Ix	IMPGS	Imports of Goods and Services
8	Zx	EXPGS	Exports of Goods and Services
9	L	HOANBS	Nonfarm Business Sector: Hours Worked for All ...
10	w	CES0500000003	Average Hourly Earnings of All Employees, Tota...
11	defl	A191RI1Q225SBEA	Gross Domestic Product: Implicit Price Deflator
12	recs	JHGDPBRINDX	GDP-Based Recession Indicator Index
13	Pop	CNP16OV	Population Level

Table 3: Descriptive statistics

	variable description	count	mean	std	min	25
Y	Gross Domestic Product	151.0	-0.0000	0.0166	-0.1137	-0.00
Y_p	Real Potential Gross Domestic Product	151.0	-0.0000	0.0022	-0.0057	-0.00
pi_s	Sticky Price Consumer Price Index less Food an...	151.0	0.0075	0.0028	0.0017	0.00
pi_c	Median Consumer Price Index (core inflation)	151.0	0.0075	0.0031	0.0003	0.00
r	Federal Funds Effective Rate	151.0	0.0082	0.0068	0.0001	0.00
I	Gross Private Domestic Investment	151.0	0.0000	0.0581	-0.2262	-0.03
C	Personal Consumption Expenditures Excluding Fo...	151.0	0.0000	0.0043	-0.0144	-0.00
Ix	Imports of Goods and Services	151.0	0.0000	0.0616	-0.3022	-0.02
Zx	Exports of Goods and Services	151.0	0.0000	0.0581	-0.3098	-0.02
L	Nonfarm Business Sector: Hours Worked for All ...	151.0	0.0000	0.0218	-0.1394	-0.01
w	Average Hourly Earnings of All Employees, Tota...	67.0	-0.0000	0.0059	-0.0135	-0.00
S	S	90.0	-0.0000	0.0240	-0.0998	-0.01