Royal Holloway, University of London

Essays on Fiscal Rule Design and its implications

Jan Fleischhacker

A thesis submitted to the Department of Economics of the university Royal Holloway, university of London, for the degree of Doctor of Philosophy

London, 01/03/2024

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the university Royal Holloway, University of London, is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 45825 words.

Abstract

The three chapters of this thesis examine different aspects of the design of fiscal rules and their implications on the effects of policy interventions and how policy reacts to the economy.

Chapter 1 focuses on the design of fiscal rules in DSGE models, which has been shown to matter crucially in identifying the effects of policy interventions and analyses two mechanically distinct components of fiscal policy rules: fiscal rule interactions and multimodality.

In a first exercise, a set of alternative fiscal rules is considered for the benchmark Leeper, Plante and Traum (2010) model, with the main design feature being across budget block (expenditure vs taxation) interactions. The models are compared using the Bayesian data density. The results show that the Leeper, Plante and Traum (2010) model is competitive in the set but may be improved by including across-budget component interaction with taxes ordered first. Mechanically, the budget component interactions trickle down to how policy interventions are financed, showing increased coordination across blocks. In the benchmark, a government consumption shock raises the federal government's expenditures, and along the path, taxation increases to bring the debt level back to the steady state. In the recursive block models, budget impacts can be temporarily purely expansionary in that expenditure increases and taxation is reduced. Combining both aspects, it seems to reflect a temporary but coordinated approach to raise output across the expenditure and taxation categories.

Secondly, I explore the role of multimodality in fiscal rules. Herbst and Schorfheide (2016) showed that fiscal parameters in the aforementioned model can become multimodal, leading to multimodal impulse responses. In essence, what that means is that fiscal policy may have varied impacts depending on the exact posterior parameter draw. For the Leeper, Plante and Traum (2010) model, I argue using graphs and demonstrate that the source of multimodality in the model is likely the structural design of the rules. Furthermore, building on the analysis in Herbst and Schorfheide (2016), I apply bi-modal regions to the highest posterior density regions as intervals tend to overestimate uncertainty of bi-modal distributions. The results show that the effects of consumption taxation shocks not only predict different scenarios depending on the mode

but also disjointed impact scenarios. In particular, for consumption taxation shocks, the average effect of a structural shock is not a particularly likely event by itself.

In Chapter 2, I explore how fiscal policy decisions relate to the business cycle and, building on that, how the effects of policy interventions may vary depending on when policy is conducted in the business cycle. To assess this, I estimate a small to medium-sized DSGE model with expressive non-linear fiscal and monetary rules using a higher-order approximation.

The estimation procedure employed in this chapter combines several existing approaches developed by Herbst and Schorfheide (2016), Jasra et al. (2010), Buchholz, Chopin and Jacob (2021) and Amisano and Tristani (2010) to trade off computation time and inference quality. The model is estimated using Sequential Monte Carlo techniques to estimate the posterior parameter distribution and particle filter techniques to estimate the likelihood. Together, the estimation procedure reduces the estimation from weeks to days by up to 94%, depending on the comparison basis.

To assess the behaviour of the effects of fiscal policy interventions, I sample impulse responses conducted along the historical data. The results present time-varying policy rules in which the effects of fiscal shocks go through deep cycles depending on the initial conditions of the economy. Among the set of fiscal instruments, government consumption goes through the most persistent cycles in its effectiveness in stimulating output. In particular, the effects of government consumption stimulus are estimated to be more effective during the financial crisis and, later, the Covid crisis, while being less effective in periods of above steady state output like the early 2000s.

Relating the effects of specific stimulating shocks to the initial conditions using regression techniques, I show that fiscal policy is more effective at stimulating output if the interest rate and debt are low. Furthermore, the effects of government consumption are estimated to be increasing in output while tax cuts are decreasing.

As a last contribution of Chapter 2, I explore how the behaviour of the central bank and government varies depending on the business cycle by analysing sampled policy rule gradients constructed on historical data. For the central bank, the results show that in phases of high output growth, the central bank puts more emphasis on controlling inflation and less on output. As the economy shifts into crisis, the central bank reduces its focus on inflation and shifts towards bringing output growth back to target. For the fiscal side, the behaviour is heavily governed by the current debt level, and, for example, during the high debt periods of the 1990s, labour taxation became increasingly responsive to debt to stabilize the budget.

Chapter 3 applies the model developed in Chapter 2 to a forecasting exercise using the DSGE-VAR framework. The analysis confirms previous results of the literature that the DSGE-VAR framework and, by extension, DSGE models are frequently useful in aiding forecasting performance for output compared to standard models. Furthermore, I show that DSGE-VAR models can help aid forecasting performance of governmental variables like government consumption and debt quite significantly. However, there seems to be no single best methodology across all data series and forecasting settings considered, similar to the results in Gürkaynak, Kısacıkoğlu and Rossi (2014). Rather, the best-performing methodology may depend on factors like sample selection, modelling framework and potentially others.

In a novel exercise, I explore the utility of a variation of the Chapter 2 model with a Zero Lower Bound constraint for forecasting. Overall, the model performs well but is not necessarily competitive with the much simpler DSGE-VAR. However, the ZLB model does show some strength in forecasting fiscal variables.

Acknowledgements

I am incredibly grateful to both my supervisors, Dr Ija Trapeznikova and Prof. Alessio Sancetta, without whom I would have never started nor come close to the finishing line, any way I look at it.

I would also like to thank my supervisor, Prof. Andrew Mountford, for advising me.

Lastly, I would like to thank the Department of Economics at Royal Holloway for awarding me with the College Studentship, which facilitated a lot of things.

Dedication

I dedicate this thesis to my parents and my girlfriend in no particular order. To my parents, for installing in me a drive to try to understand things over just learning them by heart and game-planning with me whenever times got difficult. To my girlfriend, for preventing me from going stir-crazy and always being on my team in the long battle of attrition that is a PhD.

Contents

List of Figures

List of Tables

Chapter 1

Exploring Joint Fiscal Decision Making

1.1 Introduction

The empirical literature that focuses on fiscal multipliers is not unified in terms of its estimates. Although there is common ground with how policy can interact with the economy in general, studies have found, at times, varying estimates of the effectiveness of fiscal policy.[1] To effectively aid policymakers, this variation in results needs to be narrowed down, and the field has been very successful in identifying central factors that explain this variation. Substantive research has focused on exploring how fiscal tools interact with the economy. The design questions of lump-sum taxation vs. distortionary taxation, productive vs. un-productive government consumption, monetary and fiscal interactions, and sample selection choices have been shown to matter crucially in the evaluation of fiscal policy effectiveness. But it is not limited to these. The question of how fiscal policy responds to the economy, as governed by the fiscal response functions, has been shown to explain parts of the variation of fiscal estimates.

This chapter aims to shed light on two particular components of the design of fiscal response functions. The first emphasis is on identifying which fiscal design features are important and what the consequences for the effects of fiscal policy interventions are. In this chapter, the focus is on budget component interactions, meaning within the taxation or spending category and across interactions. The second emphasis is on tracing out the source and consequences for policy analysis of a fiscal rule set with multimodal parameters in a DSGE model. Of particular interest are the consequences of uncertainty on the effects of policy interventions.

I begin the analysis with the model developed by Leeper, Plante and Traum (2010) as a benchmark for the policy design comparisons. This model forms an excellent basis for establishing

[1] Multiplier estimates in VAR applications range from around unity (Blanchard and Perotti (2002), Favero and Giavazza (2012)) up to upwards of three (Mountford and Uhlig (2009)). In VAR applications, government consumption multipliers are more stable (Caldara and Kamps (2017)). However, under special conditions like the Zero Lower Bound, government consumption can become more significantly effective (Woodford (2011), Drautzburg and Uhlig (2015) and Boubaker, Khuong Nguyen and Paltalidis (2018))

a comparative framework for several reasons. Firstly, the model features a fiscal rule set that allows the government to respond to output and debt fluctuations and also allows for interaction between tax rates. The tax rate interactions are assumed to be symmetric. In this case, symmetric implies that the effect of, for example, a capital taxation shock on labour taxation is identical to the effect a labour taxation shock would have on capital taxation. As such, it is a model with a particularly detailed description of the policy mechanism.[2] Secondly, fixing the overall model design, estimation, and data then allows for a model comparison analysis to compare different fiscal policy designs using the marginal data densities.

This chapter first draws a parallel between the structural equation modelling common to fiscal rules in DSGE models and the underlying reduced form process for the tax system in the Leeper, Plante and Traum (2010) model. Viewing the tax system in its reduced form, even in a restrictive case, allows the econometrician to think about how the restrictions placed on the structural process generate the dynamics of observables. Using this, I show that the fiscal interaction parameters in the Leeper, Plante and Traum (2010) model define long-run symmetric interactions between taxation rates. Furthermore, the multimodality observed in Herbst and Schorfheide (2016) in the Leeper, Plante and Traum (2010) model is graphically shown to be a design feature of the structural fiscal rules around the posterior estimates. Additional modes are proposed using the graphical technique and confirmed during an additional estimation.

In addition to this, I delve into the role of parameter multimodality of the model first observed in Herbst and Schorfheide (2016) using Sequential Monte Carlo (SMC) techniques. Multimodality in the posterior distributions of the model parameters occurs if the density features multiple modes and thus is not globally identified.[3] Furthermore, Herbst and Schorfheide (2016) show that multimodality in the parameter dimension may trickle down to impulse responses and policy

[2] The literature review below shows that the Leeper, Plante and Traum (2010) model features a particular rich rule set.

[3] From a frequentist perspective, this is similar to finding multiple, local modes with comparable likelihoods upon repeated runs of an optimizer. As the likelihoods are comparable, it is difficult to discard one mode over the other. In that case, the econometrician faces a not uniquely identified model with possible different properties at different modes.

analysis. In particular, they show that tax rate shocks can create multimodal impulse responses of output.

To assess the multimodality in impulse responses, I implement approaches proposed in Chen and Shao (1999) and Chen et al. (2000) for evaluating the highest posterior density intervals and bi-modal regions to apply them to impulse responses. The key idea is to show how exactly the impulse responses are affected by multimodality and to narrow down what the actually likely policy scenarios are. If one applies the standard highest posterior density intervals, one can overstate the uncertainty of the object quite significantly if the modes are well separated. The bi-modal regions, by comparison, tend to give much tighter and sometimes well-separated estimates.

The budget and stimulus dynamics of consumption taxation rate shock are affected by multimodality the most for the baseline Leeper, Plante and Traum (2010) model. Here, consumption taxation rate shocks show two types of behaviour, just as in the Herbst and Schorfheide (2016) application. The first type is behaviour typically associated with tax rates where output falls, taxation income rises, and debt falls. In addition to this, the same consumption taxation shocks seem to also imply an alternative dynamic based on this specific structural parameterization where output rises while the government deficit increases. Using the bimodal regions, I illustrate that the effects of consumption taxation, as implied by this model, predict not only different but disjointed impact scenarios. I show that on the financing side for the government, these two scenarios are well separated in the sense that the average effect of a policy intervention on government finance is not a particularly likely event in and of itself.

Next to multimodality, a different aspect that is crucial is the design of fiscal response functions. In this chapter, I conduct a non-normative analysis of fiscal policy effects and their dependence on the fiscal rule set. Here, I consider combinations of forms encountered in the literature. For within budget blocks (expenditure vs. taxation), I consider two candidates: independent or symmetric interactions, as in Leeper, Plante and Traum (2010). For across interactions, the government can set spending and taxation either independently or in a block recursive structure with either category ordered first. The result is a set of 12 alternative specifications.

Using a posterior density comparison via the Bayes factor, three key models are selected per across-interaction category. Results show that the original specification by Leeper, Plante and Traum (2010) is very robust and in the top fitting group. The only alternative specification that is preferable in terms of the data density is identical in terms of intra-category interactions but has taxes ordered first in the across-interaction structure. The impulse response analysis shows similar abilities to stimulate the economy for all three best-fitting models with minor differences for transfers. The different interaction structures seem to matter significantly for government finance. The original model predicts balanced expansionary stimulus. A government spending (taxation) shock increases overall spending (taxation income) and is accompanied by an increase in taxation income (spending), which reduces the debt impact and balances the debt expansion. The recursive block models differ here. Temporarily, the budget impacts can be purely expansionary, where both taxation income is reduced, and spending is increased. This represents a coordinated effort to influence the economy across the budget components. Consequently, moving towards a richer VAR design may allow for fiscal policy mechanisms in DSGE models to become more expressive and allow the data to speak more clearly.

On the technical side, for most estimations, the novel Sequential Monte Carlo (SMC) sampling process proposed by Herbst and Schorfheide (2016) is applied. The advantage of Sequential Monte Carlo estimations is that it uses the multinomial distribution to approximate the posterior. The approximations tend to be very capable of exploring complex posterior distribution even under non-normality and multimodality, and therefore, the SMC algorithm is uniquely suited to the problems encountered in this chapter. For detailed resources on Sequential Monte Carlo sampling, see Herbst and Schorfheide (2016), Herbst and Schorfheide (2014) and Cappé, Godsill and Moulines (2007).

Structurally, the chapter first discusses the Leeper, Plante and Traum model in section 1.3 , followed by the explorative exercise on fiscal rules in section 1.4 and multimodality in the Leeper, Plante and Traum (2010) rules. Section 1.5 provides a detailed description of the estimation proceedings. Section 1.6 explores the estimation results for the original model and the diffuse variation. Section 1.7 further discusses the role of multimodality in the model by showing the existence of further modes using slight tweaks. In addition to this, results for the estimation on a replicated data set are presented. Section 1.8 details the main results of the chapter on a diverse set of interaction structures. Then, section 1.9 reapplies the preferred model and engages in robustness analysis on the extended data set.

1.2 Literature review

The way fiscal policy rules are most commonly constructed in DSGE models is by treating the government as a simplified decision-maker. This decision-maker sets the current level of a group of fiscal instruments, which may include taxation and expenditure variables or may include choosing a debt level according to prespecified structural rules and a budget constraint. Unlike households and firms, the federal government commonly does not optimize a target but follows a set rule. Hence, designing good fiscal rules is a key component to understanding fiscal policy and the effects that fiscal policy has on the economy. So, in a sense, fiscal policy in DSGE models is mostly descriptive and seeks to answer what the government does and what its effects are.

In the federal government decision process, one may distinguish between two types of variables. The first type is one that “forcibly” adjusts to close the budget. For example, let’s assume a government has chosen a current taxation and expenditure level. Unless the current choices perfectly close the budget, a final variable may need to adjust to ensure the budget constraint holds. This variable is typically chosen to be either debt or lump sum taxes. The second type of variable is given a prescribed structural rule on how it’s meant to be set. Within this structural rule, the fiscal variables may respond to output fluctuations or to other variables in the government’s information set. Typically, the models assume a linear fiscal rule so that the marginal responses are constant across the business cycle.[4]

With the option to specify structural rules comes a lot of design freedom for economists. The typical modelling starts at independent AR(1) processes for the individual instruments. For example, in the Leeper, Plante and Traum (2010) model, the consumption taxation rate is treated as an AR(1) process, where the current rate only depends on itself one quarter ago and an

[4] Chapter 2 explores a type of fiscal rule which responds differently depending on the business cycle. The resulting federal government goes through phases in the way that it responds to the economy. These phases are induced based on the business cycle conditions.

exogenous shock. No further factors go into determining the current rate. Fiscal rules can extend from thereon.

Most extensions can generally be grouped into a set of restricted linear state space models of varying complexity and interactions. One way fiscal rules can be extended is to include interactions with the economy and other factors of interest. That typically covers the interactions of fiscal variables with same-period output and public debt levels but can also include other variables.[5] Studies by Zubairy (2014) and Leeper, Plante and Traum (2010) have delved into the importance of how fiscal rules respond to federal debt or output. The former study conducts a counterfactual analysis for a DSGE model with fiscal rules where taxation rates and government consumption are allowed to respond to both output and debt steady state deviations. In the analysis, the speed of adjustment to either debt or output is varied using a scaling factor. They show that fiscal multipliers become less impactful for very high levels of debt responsiveness. In addition, if fiscal instruments react more aggressively to output steady state deviations, then the effects of government consumption shocks are more muted. The latter study explores the effects of fiscal shocks on output via a similar counterfactual experiment using a variable fiscal adjustment speed for federal debt. Leeper, Plante and Traum (2010) find that for higher adjustment speeds to debt, government consumption multipliers are smaller. The clear result is that the design of the exogenous structure matters.

The other set of typical modifications focuses on the autoregressive matrices of the state space model. This is also what this chapter is concerned with. The simplest way one may restrict a model is to assume that the individual variables follow independent AR(1) processes. Fully independent tax and expenditures structures are very popular and are frequently implemented (e.g. see Mucka and Horvath (2022), Bondzie and Armah (2022), Fernández-Villaverde et al. (2015), Zubairy (2014), Drautzburg and Uhlig (2015) and more). Departing from the independent case allows economists to imbue DSGE governments with a richer and more complex ruleset with interactions both within fiscal taxation and spending budget components but also across. Leeper, Plante and Traum (2010), Traum and Yang (2015) and Yang (2005) all consider symmetric tax

[5] For an interesting case, Chan (2020) allows fiscal variables to adjust to emission levels.

rules but don't allow for direct interactions between spending and taxation. The role of symmetric rules is interesting as they can act in budget-balancing ways and imply a high degree of coordination on the policymaker's side. Across fiscal block interactions, as implemented in Davig and Leeper (2011) in a block-recursive format, imply a sequential decision process. First, the policymaker decides on a government spending (taxation) level. In the second and final step, the government adjusts its taxation (spending) variables and debt levels to close the budget.

To get an understanding of the current design of fiscal policy rules in papers on fiscal policy, I surveyed 30 papers and categorized their fiscal rules for tax rates and government consumption. I chose to focus on the most recent papers published since 2015 in an academic journal, a central bank working paper series, or similar.[6]

For each model, only time-varying tax components were assessed. Fixed components, such as a fixed income tax rate, were omitted as not relevant. The tax structure was then described based on three components (Lump sum taxation, single marginal tax rate and multiple marginal tax rates) and, possibly, their combinations. To assess the design of the fiscal policy rules in the specific papers, I avoid categorizing all rules and focus on the most elaborate rule. For the tax rates, the most elaborate rule is categorized by (a) if it features an autoregressive component[7], (b) if it responds to output and debt and (c) if so, to which variables[8]. Similarly, for government consumption, the rule is characterized by if it features an autoregressive component and whether it responds to output and debt. The last feature of interest was whether, in the feedback rules, the model allowed for interactions between tax rates and government consumption.

[6] I used google scholar with the search term "DSGE "fiscal policy"" to search for papers. Further criterions were that the paper was written in English, was accessible, featured a DSGE model and fiscal rules were defined on at least one of the two categories: tax rules or government consumption. Additional fiscal components were not surveyed. The sample has an average publishing years of 2018 and 32 citations on average.

[7] Autoregressive features can enter in several ways. For example, the tax rate might linearly depend on its past values, but it also may depend non-linearly on past values in the structural equation. Equally, the tax rate may depend on a structural shock that is governed by an AR(p) process.

[8] Tax instruments may depend linearly or non-linearly on debt and output in the structural equations.

Table 1.1: Review of Survey Results on fiscal rule design

Paper	**Tax**			**Government Cons.**		**Interactions**
	Type	AR	B and/or Y	AR	B and/or Y	
Shen and Yang (2016)	single marginal	Yes	B	Yes	Y	No
Bhattarai and Trzeciakiewicz (2017)	multiple marginal	Yes	B,Y	Yes	B,Y	No
Mucka and Horvath (2022)	multiple marginal	Yes	No	Yes	No	No
Takyi and Leon-Gonzalez (2020)	multiple marginal	Yes	B,Y	Yes	Y	No
Samimi et al. (2017)	multiple marginal	Yes	B,Y	Yes	B,Y	No
Drautzburg (2020)	multiple marginal	Yes	B,Y	Yes	B,Y	No
Xiao, Fan and Guo (2018)	multiple marginal	Yes	No	Yes	No	No
Mumtaz and Theodoridis (2020)	multiple marginal	Yes	B,Y	Yes	B,Y	No
Chan (2020)	Lump sum, single marginal	Yes	No	Yes	No	No
Wang (2021)	Lump sum, multiple marginal	No	B,Y	Yes	No	No
Aursland et al. (2020)	Lump sum, multiple marginal	Yes	No	Yes	No	No
Cavalcanti and Vereda (2015)	Lump sum, multiple marginal	Yes	No	Yes	No	No
Drygalla, Holtemöller and Kiesel (2018)	Lump sum, multiple marginal	Yes	B,Y	Yes	B,Y	No
Babecký, Franta and Ryšánek (2018)	Lump sum, multiple marginal	Yes	B,Y	Yes	B,Y	Yes
Carvalho and Castro (2017)	Lump sum, multiple marginal	Yes	No	Yes	B	No
Bušs and Grüning (2023)	Lump sum, multiple marginal	Yes	B,Y	Yes	B,Y	No
Bondzie and Armah (2022)	Lump sum, multiple marginal	Yes	No	Yes	No	No
Gadatsch, Hauzenberger and Stähler (2016)	Lump sum, multiple marginal	Yes	B,Y	Yes	B,Y	No
Gomes et al. (2015)	Lump sum	No	B	Yes	No	No
Fève and Sahuc (2016)	Lump sum	No	No	Yes	No	No
Li and Spencer (2015)	Lump sum	No	No	Yes	Y	No
Zeman (2017)	Lump sum	No	B,Y	No	No	No
Faria-e-Castro (2021)	Lump sum	No	B	No	No	No
Engler and Tervala (2018)	Lump sum	No	No	Yes	No	No
Wang et al. (2020)	Lump sum	Yes	No	No	No	No
Kliem, Krowlusky and Sarferaz (2016)	Lump sum	Yes	B	Yes	B	No
Kollman et al. (2016)	Lump sum	Yes	B,Y	Yes	Y	No
Jesus, Besarria and Maisa (2020)	--	--	--	Yes	No	No
Andreyev (2020)	--	--	--	Yes	Y	No
Kang and Suh (2017)	--	--	--	Yes	No	No

Notes: The survey results table presents summaries of fiscal rules in the respective papers. Results are separated into tax and government consumption columns. In the second column, the type of tax instruments in the papers is classified as Lump sum, marginal and combinations thereof. The third column gives a simple "Yes" or "No" indicator if the rule includes an autoregressive component, and the fourth column shows whether the fiscal instrument responds to debt (B) and/or output (Y). Papers that only include fixed tax rules are indicated as missing. Government consumption rules are summarized in a similar way. The last column includes a "yes" or "no" indicator for whether there are interactions between the fiscal blocks.

Table 1.1 presents the results of the survey. While the literature review presented here is by no means exhaustive, it helps to identify a number of common features. For the tax components, the most common type of fiscal rule for the most elaborate tax component features an autoregressive component with 74%, and about 59% of tax rules respond to at least output or debt (or both). These figures exclude models with fixed tax components. On the flip side, that implies that a substantial share of rules is very simple in design, with 26% of rules not featuring autoregressive components that induce persistence in taxation and about 41% of tax rules do not respond to debt or output to balance the budget. For the government consumption rules, 90% of rules include an autoregressive component, while 50% of papers allow government consumption to respond to output or debt (or both). Overall, the majority of papers surveyed allow fiscal instruments to respond to past values and the economy.

The next focus is on whether taxes and spending respond to each other. If the government wishes to increase government consumption today, it seems reasonable that the government may consider adjusting tax instruments in tandem with changes in spending. Interestingly, only one paper out of thirty papers included interaction between the two fiscal blocks in the feedback rules. Babecký, Franta and Ryšánek (2018) includes interaction between all instruments based on the structural shock processes, and thus, their model allows for the direct response of one fiscal block to the other. In part, they base their interaction design on the work by Leeper, Plante and Traum (2010). Based on this, further analysis into richer fiscal rulesets, as it is conducted here, maybe a useful contribution to the literature.

1.3 Leeper, Plante and Traum (2010) model

The following section will introduce the reader to this paper's core model developed by Leeper, Plante and Traum (2010). This model is a small to medium-sized neoclassical growth model with key fiscal features such as distortionary taxation, a rich set of fiscal feedback rules and fiscal financing design. In addition, other commonly used fiscal policy rules can be considered as a special case of the Leeper, Plante and Traum (2010) model with additional restrictions.

1.3.1 The household problem

The core of the household problem is a constrained dynamic programming problem, where the household optimises the sum of the discounted path of utility. To be precise, the agent maximises:

$$\max_{c_t, l_t} \qquad E_0 \sum_{t=0}^{\infty} \beta^t \, u_t^b \left[\frac{(\,c_t - hC_{t-1})^{1-\gamma}}{1-\gamma} - u_t^l \, \frac{l_t^{1+\kappa}}{1+\kappa} \right],$$

subject to the budget constraint and capital transition equation:

$$(1+\tau_t^C)c_t + i_t + b_t = (1-\tau_t^K)R_t^K v_t k_{t-1} + (1-\tau_t^L)l_t w_t + R_{t-1}b_{t-1} + z_t$$

$$k_t = \big(1-\delta(v_t)\big)k_{t-1} + \left[1 - s\left(\frac{u_t^i i_t}{i_{t-1}}\right)\right] * i_t,$$

where capitalized variables refer to aggregate variables, and lowercase variables denote individual variables. The agent derives utility from consumption, c_t. c_t is seen relative to the previous period's habit-adjusted aggregate consumption level, hC_{t-1}. Here, the habit parameter, h, is assumed to be in the open interval from zero to one ($h \in (0,1)$). The marginal benefit is governed by the risk aversion parameter $\gamma > 0$, which ensures that the marginal benefit of consumption for finite levels is always positive but decreasing in consumption. To finance spending, the agent can supply labour hours, l_t, to the firm in return for labour income, $l_t w_t$, based on the wage rate, w_t. Simultaneously, the household faces a utility cost of working, which is additively separable to consumption. Together, the two parts of the utility are distorted by two sources. Firstly, a preference shock, u_t^b, distorts both parts together. Secondly, the labour supply shock, u_t^l, directly impacts the marginal disutility of labour.

The household can hold two assets to smooth consumption and accumulate wealth. The first asset is a one-period government debt bond, b_t, that pays an interest R_t in the next period. The second important asset is the capital stock, k_t. The capital stock today is determined in a standard fashion based on yesterday's stock minus its depreciation and plus the newly invested capital, i_t.

However, the capital law of motion includes two special features. Firstly, capital depreciation is not constant, but instead, the depreciation rate, δ, is governed by the following function:

$$\delta(v_t) = \delta_0 + \delta_1(v_t - 1) + \frac{\delta_2}{2}(v_t - 1)^2.$$

The depreciation rate is a function of the capital utilisation level, v_t. A higher utilisation level augments firm production but increases the capital stock's depreciation. δ_0 corresponds to the steady state level of the depreciation rate. The second feature is found in the adjustment cost for varying the investment level, $s\left(\frac{u_t^i i_t}{i_{t-1}}\right)$. The adjustment cost depends on the ratio of current to previous investment and is perturbed by the adjustment cost shock, u_t^i. At the steady state, the function s satisfies the following properties: $s(1) = s'(1) = 0$ and $s''(1) > 0$. These properties define the value of the adjustment function and its gradient at the steady state of the system as $\frac{u_t^i i_t}{i_{t-1}}|_{ss} = 1$. For a linear DSGE estimation as in Leeper, Plante and Traum (2010), one only requires the level of $s\left(\frac{u_t^i i_t}{i_{t-1}}\right)$ at the steady state to determine the remaining steady state values and the gradient of $s\left(\frac{u_t^i i_t}{i_{t-1}}\right)$ at the steady state to solve the system. No further functional assumptions are needed. With this penalisation, the consumer prefers gradual adjustments of the investment level to avoid higher adjustment costs. This generates some stability in the investment series, as seen in the data. The laws of motion of shocks in the household problem are governed by AR(1) processes:

$$ln(u_t^b) = \rho_b ln(u_{t-1}^b) + \sigma_b \varepsilon_t^b, \qquad \varepsilon_t^b \sim N(0,1),$$

$$ln(u_t^l) = \rho_l ln(u_{t-1}^l) + \sigma_l \varepsilon_t^l, \qquad \varepsilon_t^l \sim N(0,1),$$

$$ln(u_t^i) = \rho_i ln(u_{t-1}^i) + \sigma_i \varepsilon_t^i, \qquad \varepsilon_t^i \sim N(0,1).$$

These are defined in terms of the natural logs of the shock. Specifying the laws of motion in terms of logs ensures that the shocks are constrained to be larger than 0. Therefore, the shocks are consistent with the phrasing of the optimisation problem above (i.e. $u_t^l < 0$ would imply that working more generates utility).

The federal government interacts with the household in several ways via its budget constraint. Firstly, consumption, labour income and the return on capital investment are all taxed via distortionary taxation rates: τ_t^C, τ_t^L and τ_t^K. Here, τ_t^C is the consumption tax rate, τ_t^L is the labour tax rate and τ_t^K is the capital tax rate. Furthermore, the household may receive varying levels of transfer payments, z_t.

1.3.2 The firm problem

Firms face a well-studied production problem. The representative firm maximises within-period profits by choosing capital and labour levels. Together, capital and labour are used to produce output through a Cobb-Douglas production function. The production function is subject to the capital utilisation level, v_t, selected by the consumer and a technology shock process, u_t^a. At the same time, the firm faces cost for its capital and labour utilisation. The labour cost is based on the wage rate, w_t, and the capital cost depends on the rental rate, R_t^K, and the utilization level, v_t. The formulation of the problem is as follows:

$$\max_{k_{t-1},l_t} u_t^a (v_t k_{t-1})^\alpha (l_t)^{1-\alpha} - w_t l_t - R_t^K v_t k_{t-1}.$$

The productivity shock, u_t^a, is driven by a AR(1) process defined on the log values:

$$ln(u_t^a) = \rho_a ln(u_{t-1}^a) + \sigma_a \varepsilon_t^a, \qquad \varepsilon_t^a \sim N(0{,}1).$$

1.3.3 The government problem

The last remaining component of the DSGE model is the federal government. The federal government responds to aggregate variables that summarize the economy, and variables like consumption and labour hours are capitalized within the federal government problem. The federal government faces the following constraint to ensure solvency:

$$B_t + \tau_t^K R_t^K v_t K_{t-1} + \tau_t^L w_t l_t + \tau_t^C C_t = B_{t-1} R_{t-1} + G_t + Z_t.$$

The federal government can sell one-period bonds, B_t, to finance its operations if necessary but pays an interest rate, R_t, on it. Further, the government receives taxation income based on the distortionary taxation instruments τ_t^K, τ_t^L and τ_t^C and the corresponding tax bases. Furthermore, it has expenditures in the form of government consumption, G_t, and transfers, Z_t.

This model features an overall neutral description of fiscal policy. As opposed to the household or firm problem, the government rules are not chosen based on an optimisation problem. Instead, the focus is on entertaining a rich but neutral linear ruleset. The ruleset is defined on the fiscal terms in terms of log steady state deviations indicated by the hatted variables. Looking at the government expenditure side, government consumption, $\widehat{G}_t$, and transfers, $\widehat{Z}_t$, respond to current economic development in terms of $\widehat{Y}_t$ and also the previous debt level, $\widehat{B}_{t-1}$. The prior space for the fiscal rules implies that the government reduces spending in phases of above steady state output and debt. Thus, it is budget balancing. Both spending categories obtain some level of persistence via structural autoregressive shocks.

$$\widehat{G}_t = -\varphi_g \widehat{Y}_t - \gamma_g \widehat{B}_{t-1} + \hat{u}_t^g,$$

$$\hat{Z}_t = -\varphi_z \hat{Y}_t - \gamma_z \widehat{B}_{t-1} + \hat{u}_t^z,$$

$$\hat{u}_t^g = \rho_g \hat{u}_{t-1}^g + \sigma_g \varepsilon_t^g, \qquad \varepsilon_t^g \sim N(0{,}1),$$

$$\hat{u}_t^z = \rho_z \hat{u}_{t-1}^z + \sigma_z \varepsilon_t^z, \qquad \varepsilon_t^z \sim N(0{,}1).$$

The taxation processes are constructed in a similar fashion. Capital and labour taxation can change in response to rising output and debt. Again, the design has a budget-balancing character. The only variable exempt from this is consumption taxation. The reason is that on the federal level, consumption taxes are used to target specific goods. An additional feature of the Leeper, Plante and Traum (2010) ruleset is the symmetric interaction between tax rates. This implies, for example, that a unit shock to the capital tax rate has the same effect on the labour tax rates as a labour tax shock has on the capital tax rate (captured by ϕ_{kl}). The rules for tax rates and their corresponding structural shocks are as follows:

$$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \widehat{B}_{t-1} + \phi_{kl} \hat{u}_t^{\tau^l} + \phi_{kc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^k},$$

$$\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \widehat{B}_{t-1} + \phi_{kl} \hat{u}_t^{\tau^k} + \phi_{lc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^l},$$

$$\hat{\tau}_t^c = \phi_{kc} \hat{u}_t^{\tau^k} + \phi_{lc} \hat{u}_t^{\tau^l} + \hat{u}_t^{\tau^c},$$

$$\hat{u}_t^{\tau^k} = \rho_{\tau^k} \hat{u}_{t-1}^{\tau^k} + \sigma_{\tau^k} \varepsilon_t^{\tau^k}, \qquad \varepsilon_t^{\tau^k} \sim N(0{,}1),$$

$$\hat{u}_t^{\tau^l} = \rho_{\tau^l} \hat{u}_{t-1}^{\tau^l} + \sigma_{\tau^l} \varepsilon_t^{\tau^l}, \qquad \varepsilon_t^{\tau^l} \sim N(0{,}1),$$

$$\hat{u}_t^{\tau^c} = \rho_{\tau^c} \hat{u}_{t-1}^{\tau^c} + \sigma_{\tau^c} \varepsilon_t^{\tau^c}, \qquad \varepsilon_t^{\tau^c} \sim N(0{,}1).$$

All parameters in the above equations, including autoregressive parameters, interaction terms and standard deviations, are fully estimated once the model is taken to the data.

1.3.4 Market clearing and model solution

In equilibrium, all goods produced need to be consumed in some fashion. In this model, the aggregate constraint is as follows:

$$Y_t = C_t + I_t + G_t.$$

Together with the closing condition, the complete model is defined by the household's first-order conditions, capital accumulation law, first-order conditions of the firm problem, government budget constraint and the various autoregressive rules for fiscal variables and structural shocks. The model is solved using the procedure developed in Sims (2002).

1.4 Exploring tax systems and structural rules

1.4.1 The reduced form fiscal system and impulse responses

To gain a deeper understanding of the assumptions placed on the tax rules by Leeper, Plante and Traum (2010), this section solves the structural fiscal rules to a standard VAR rule under the assumption that the economy is exogenous. Based on this form and the corresponding system matrices, one can see that the symmetry assumption placed on the tax rules holds for impulse responses on impact but also at longer horizons. Furthermore, one can explicitly see how the structural parameters relate to the VAR parameters. The following sections exploits this by showing that the multimodality found by Herbst und Schorfheide (2016) likely arises due to the non-linear relationship between structural parameters and VAR parameters.

The tax system in Leeper, Plante and Traum (2010) has three major components. Firstly, the exogenous component, X, is modelled using linear responses of the capital and labour tax rates to current output, $\varphi_i \hat{Y}_t$, and previous debt, $\gamma_i \widehat{B}_{t-1}$, in terms of steady state deviations. The second essential component is the laws of motion for the structural shocks $\hat{u}_t^{\tau^k}$, $\hat{u}_t^{\tau^l}$ and $\hat{u}_t^{\tau^c}$. These are all modelled as independent autoregressive processes of order one. The last component defines the interaction between tax rates and structural shocks. Here, Leeper, Plante and Traum (2010) opted for a symmetric scheme. The assumption is that, for example, a capital taxation shock has the same effect on the labour taxation rate, as a labour taxation shock of the same size would have on the capital taxation rate. The tax system is governed by the following set of equations:

$$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \widehat{B}_{t-1} + \phi_{kl} \hat{u}_t^{\tau^l} + \phi_{kc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^k},$$

$$\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \widehat{B}_{t-1} + \phi_{kl} \hat{u}_t^{\tau^k} + \phi_{lc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^l},$$

$$\hat{\tau}_t^c = \phi_{kc}\hat{u}_t^{\tau^k} + \phi_{lc}\hat{u}_t^{\tau^l} + \hat{u}_t^{\tau^c},$$

$$\hat{u}_t^{\tau^k} = \rho_{\tau^k}\hat{u}_{t-1}^{\tau^k} + \sigma_{\tau^k}\varepsilon_t^{\tau^k}, \qquad \varepsilon_t^{\tau^k} \sim N(0,1),$$

$$\hat{u}_t^{\tau^l} = \rho_{\tau^l}\hat{u}_{t-1}^{\tau^l} + \sigma_{\tau^l}\varepsilon_t^{\tau^l}, \qquad \varepsilon_t^{\tau^l} \sim N(0,1),$$

$$\hat{u}_t^{\tau^c} = \rho_{\tau^c}\hat{u}_{t-1}^{\tau^c} + \sigma_{\tau^c}\varepsilon_t^{\tau^c}, \qquad \varepsilon_t^{\tau^c} \sim N(0,1).$$

This system can be rewritten into the following structural linear state space system using matrix notation:

$$\tau_{\mathrm{t}} = \Gamma_0 \mathrm{z_t} + \Gamma_1 \mathrm{u_t} \qquad \text{and} \qquad \mathrm{u_t} = \Lambda \mathrm{u_{t-1}} + \varepsilon_{\mathrm{t}}, \quad \varepsilon_{\mathrm{t}} \sim N(0, \Theta),$$

where $\tau_{\mathrm{t}} = [\hat{\tau}_t^k, \hat{\tau}_t^{\mathrm{l}}, \hat{\tau}_t^{\mathrm{c}}]'$, $\mathrm{z_t} = [\hat{Y}_t, \widehat{B}_{t-1}]'$ and $\mathrm{u_t} = [\hat{u}_t^{\mathrm{k}}, \hat{u}_t^{\mathrm{l}}, \hat{u}_t^{\mathrm{c}}]'$. Θ corresponds to the diagonal Covariance matrix of the structural shocks based on the three standard deviation parameters σ_{τ^k}, σ_{τ^l} and σ_{τ^c}. Γ_1 is a symmetric matrix with diagonal entries equal to 1 and is generated from $\theta = [\phi_{kl}, \phi_{\mathrm{kc}}, \phi_{\mathrm{lc}}]'$. Γ_0 governs the responses of the tax rates to output and government debt:

$$\Gamma_1 = \begin{bmatrix} 1 & \phi_{kl} & \phi_{kc} \\ \phi_{kl} & 1 & \phi_{\mathrm{lc}} \\ \phi_{kc} & \phi_{\mathrm{lc}} & 1 \end{bmatrix}, \qquad \Gamma_0 = \begin{bmatrix} \varphi_k & \gamma_k \\ \varphi_l & \gamma_l \\ 0 & 0 \end{bmatrix}.$$

Γ_1 is of full rank if $\det(\Gamma_1) = 1 + \phi_{\mathrm{kc}}^2 + \phi_{\mathrm{lc}}^2 + \phi_{\mathrm{kl}}^2 + 2\phi_{kc}\phi_{\mathrm{lc}}\phi_{\mathrm{kl}} \neq 0$. As the number of structural shocks in $\mathrm{u_t}$ is equal to the number of tax variables, one can solve for the reduced form VARX. Assuming invertibility, we can derive the underlying reduced form vector process of τ_{t} as follows:

$$\tau_{\mathrm{t}} = \Gamma_0 \mathrm{z_t} + \Gamma_1 \Lambda \Gamma_1^{-1}(\tau_{\mathrm{t-1}} - \Gamma_0 \mathrm{z_{t-1}}) + \Gamma_1 \varepsilon_{\mathrm{t}}, \;\; \varepsilon_{\mathrm{t}} \;\sim N(0, \Theta),$$

$$\tau_{\mathrm{t}} = A\mathrm{x_t} + \mathrm{B}\tau_{\mathrm{t-1}} + v_{\mathrm{t}}, \quad v_{\mathrm{t}} = \Gamma_1 \varepsilon_{\mathrm{t}} \sim N(0, \Gamma_1 \Theta \Gamma_1') = N(0, \Omega).$$

$\mathrm{x_t}$ stacks the exogenous state variable together into one vector such that $\mathrm{x_t} = [\mathrm{z_t}, \mathrm{z_{t-1}}]'$. Further, $B = \Gamma_1 \Lambda \Gamma_1^{-1}$ and $A = [\Gamma_0, \text{-}\, \Gamma_1 \Lambda \Gamma_1^{-1} \Gamma_0] = [\Gamma_0, \text{-}\, B\Gamma_0]$. Comparing the connection between the structural and reduced-form approaches, one can see how the former's parameters generate the latter in the reduced-form model. The AR(1) coefficient matrix, B, is generated via directly parameterising its eigenvalue decomposition, $\Gamma_1 \Lambda \Gamma_1^{-1}$. The autoregressive parameters of the processes, ρ_{τ^k}, ρ_{τ^l} and ρ_{τ^c}, are the eigenvalues of B. As such, they generate the persistence of any shocks. The column vectors of Γ_1, $q_1 = [1, \phi_{kl}, \phi_{kc}]$, $q_2 = [\phi_{kl}, 1, \phi_{lc}]$ and $q_3 = [\phi_{kc}, \phi_{lc}, 1]$ are eigenvectors of B. The eigenvectors are points in the state space for which τ_{t} is only stretched by the eigenvalues, ignoring shocks and the exogenous component. Additionally, Γ_1 also influences the covariance matrix of the reduced form shock process, Ω. Putting things together, the impulse response of the tax rates to shocks in this isolated system can be defined by:

$$IRF(\tau_{\mathrm{t+s}}|v_{\mathrm{t}}) = E(\tau_{\mathrm{t+s}}|v_{\mathrm{t}}) - E(\tau_{\mathrm{t+s}}|v_{\mathrm{t}} = 0) = B^s v_{\mathrm{t}} = (\Gamma_1 \Lambda \Gamma_1^{-1})^s \Gamma_1 \varepsilon_{\mathrm{t}} = \Gamma_1 \Lambda^s \varepsilon_{\mathrm{t}}.$$

Consequently, the long-run impulse response maintains the symmetric character of the structural form in this isolated system.

1.4.2 Identification

Herbst und Schorfheide (2016) showed that under less restrictive priors, the fiscal interaction parameters, $\theta = [\phi_{kl}, \phi_{\mathrm{kc}}, \phi_{\mathrm{lc}}]'$, in the Leeper, Plante and Traum (2010) model become multimodal and are not uniquely identified. Identification problems can occur when the sampler cannot distinguish between parameterisations purely based on the likelihood. There are several reasons why this might happen. Firstly, a DSGE model can include parameters that simply do not enter the likelihood in any meaningful way. In this case, all parameter choices generate the same likelihood. Secondly, non-identifiability can also occur when the likelihood is informative but does not uniquely identify a structural model based on its reduced form. The latter case seems to apply to the Herbst and Schorfheide (2016) estimation, as parameters do diverge quite significantly from the priors, and the multimodality is fairly localized. However, the question is, does the multimodality arise because the data identifies multiple laws of motion with similar likelihoods that are mechanically distinct or do the different modes generate observationally similar laws of motion? This section will attempt to illustrate that the multimodality in the Leeper, Plante and Traum (2010) model can arise due to the type of structural parameterisation of the fiscal rules. To do so, I am first going to graphically explore the type of identification issue found in Herbst and Schorfheide (2016) around their posterior estimates. As a second step, the analysis is extended to explore if the identification issue is unique to the cases found in Herbst and Schorfheide (2016).

To approach this, I separate the tax system from the rest of the DSGE model, as in the previous section and assume that debt and output are at their steady state. The system then reduces to the following VAR:

$$\tau_{\mathrm{t}} = \mathrm{B}\tau_{\mathrm{t-1}} + \upsilon_{\mathrm{t}}.$$

This system ignores the channel from and to $\mathrm{z_t}$ and $\mathrm{z_{t-1}}$ but doing so facilitates the analysis. In general, for a multimodal posterior, a requirement is that one can characterise separated θ's for which Ω and B are generated that are fundamentally similar. If the reduced form system is similar, then the estimated likelihood will be similar. Consequently, in these areas of the parameter space, the sampler cannot distinguish between different θ's and the estimation may become multimodal.

In the Herbst and Schorfheide (2016) case, there are two important bits of information about the design of the system that will be utilized here. Firstly, the only matrices that are constructed based on $\theta = [\phi_{kl}, \phi_{\mathrm{kc}}, \phi_{\mathrm{lc}}]$ is Γ_1 in $B = \Gamma(\theta)_1 \Lambda \Gamma(\theta)_1^{-1}$ and $\Omega = \Gamma(\theta)_1 \Theta \Gamma(\theta)_1{}'$. Secondly, the multimodality of ϕ_{lc} and ϕ_{kc} shown in Herbst and Schorfheide (2016) is restricted. Positive realisations of ϕ_{kc} are associated with negative values of ϕ_{lc} and vice versa. ϕ_{kl} is estimated to be unimodal and centred at around 1.57. To proceed, I choose the mode with the largest amount of posterior mass as a reference value:

$$\theta^{ref} = \left[\theta_1^{ref}, \theta_2^{ref}, \theta_3^{ref}\right] = [\phi_{kl}^{mode}, \phi_{kc}^{mode}, \phi_{lc}^{mode}] = [1.57, -3{,}1].$$

As a comparison value, I construct the parameter vector θ^* as follows:

$$\theta^* = [\theta_1^*, \theta_2^*, \theta_3^*] = [1.57, \theta_2^*, \theta_3^*].$$

For the comparison vector, I fix ϕ_{kl} to the posterior mean estimate, as this parameter is estimated to be unimodal. The remaining two parameters are allowed to vary. To compare the reduced-form systems based on θ^{ref} and θ^*, one can construct two alternative differences Δ_B and Δ_Ω as follows:

$$\Delta_B = \mathrm{B}(\theta^*) - \mathrm{B}(\theta^{ref}) = \Gamma(\theta^*)_1 \Lambda \Gamma(\theta^*)_1^{-1} - \Gamma(\theta^{ref})_1 \Lambda \Gamma(\theta^{ref})_1^{-1},$$

$$\Delta_\Omega = \Omega\ (\theta^*) - \Omega\ (\theta^{ref}) = \Gamma(\theta^*)_1 \Theta \Gamma(\theta^*)_1' - \Gamma(\theta^{ref})_1 \Theta \Gamma(\theta^{ref})_1'.$$

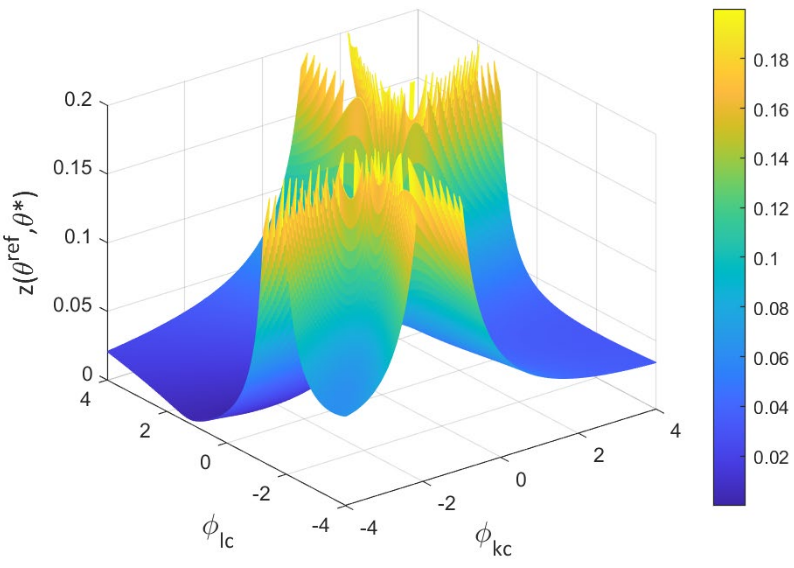
The matrices Δ_B and Δ_Ω gives us a measure of the similarity of the reduced form matrices. If, in both cases, all entries are equal to zero, then the reduced form system generated from θ and θ^* will be identical. In general, however, Δ_B and Δ_Ω may not be exactly equal to zero, but perhaps close to zero. To get an overall measure of the similarity of the reduced form matrices, I rely on the Frobenius norm applied to Δ_B and Δ_Ω . The Frobenius norm is defined as:

$$||A||_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}^2 \right),$$

where A is a n by n sized matrix. The choice of norm is not unique in this case, and different norms put a different emphasis on the comparison between the matrices. I experimented with alternatives like comparing the maximum eigenvalues. The results were similar.

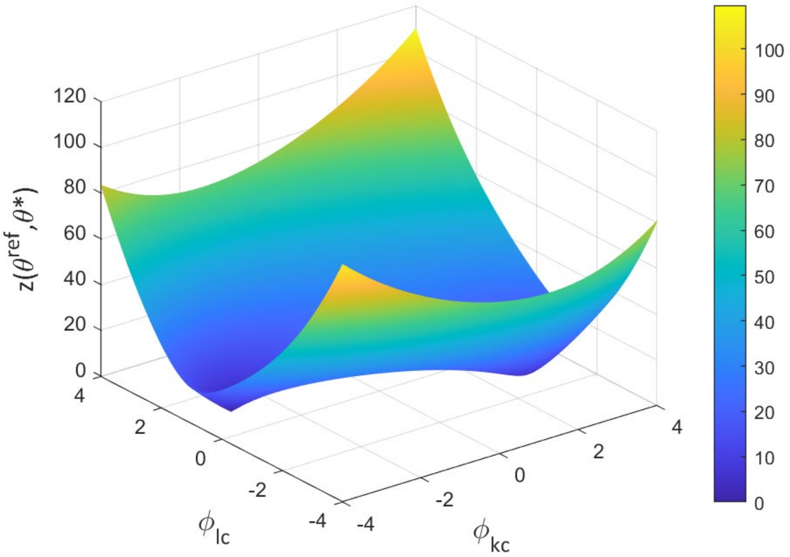
To get a sense of the space, I trace out the norm for Δ_B and Δ_Ω by varying the free parameters in θ^* across a grid. The results for Δ_B and Δ_Ω are plotted in Fig. 1.1 and Fig. 1.2 below. On the z-axis is the norm of Δ_B and Δ_Ω across the two graphs and on the x and y axis are the values for ϕ_{lc} and ϕ_{kc}.

Fig. 1.1: 3d Graph of normed distance between AR matrices based on original mode



Notes: 3d Graph plotting the normed difference, $z(\theta^{ref}, \theta^*)$, between the autoregressive matrices, $B(\theta^*)$ and $B(\theta^{ref})$, on the z-axis over a grid for ϕ_{lc} and ϕ_{kc}. θ^{ref} is set to $[1.57, -3{,}1]$ and in θ^* the first entry is fixed to 1.57.

Fig. 1.2: 3d Graph of normed distance between Covariance matrices based on original mode



Notes: 3d Graph plotting the normed difference, $z(\theta^{ref},\theta^{*})$, between the covariance matrices, $\Omega(\theta^{*})$ and $\Omega(\theta^{ref})$, over a grid for ϕ_{lc} and ϕ_{kc}. θ^{ref} is set to $[1.57,-3,1]$ and in θ^{*} the first entry is fixed to 1.57.

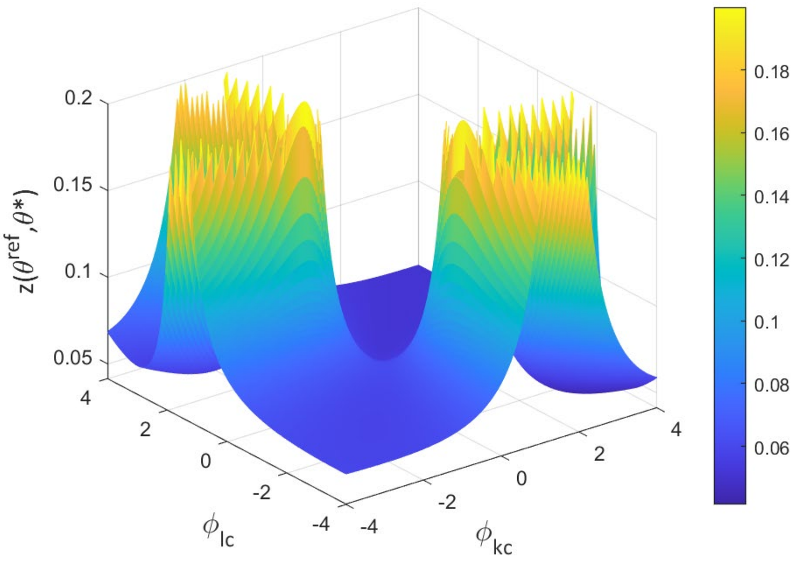
At and around the reference mode, $\theta^{ref} = [1.57, -3{,}1]$, the normed distances are naturally quite close to zero for both matrices that make up the reduced form system. However, there is an additional point in the parameter space that corresponds to the alternative mode found by Herbst and Schorfheide (2016), where the normed distance is small. That is around the point $\theta^{*} = [1.57{,}3, -1]$. What that means is that at the comparison parameter vector, θ^{*}, the constructed reduced-form system is very similar to the one at the reference point. In addition, as the autoregressive matrices are observationally similar, the matrix A that governs how the tax rates respond to output and debt ought to be very similar as well as A is influenced by the interaction parameters only through the autoregressive matrix ($A = [\Gamma_0$,- $B\Gamma_0]$). Consequently, based on the design of symmetric interaction rules for the tax rates, it can be argued that this type of system can deliver observationally similar laws of motion for the tax rates around the posterior estimates found by Herbst and Schorfheide (2016). This is problematic in the sense that there are two similarly plausible but different structural shock series around the posterior that exist because of the design of the system.

Because of the symmetric nature of the fiscal rules, there is arguably nothing particularly special about the multimodality found for ϕ_{lc} and ϕ_{kc} in comparison to the unimodality for ϕ_{kl}. It stands to reason that there may be additional unexplored modes around $\phi_{kl} = -1.57$. That is especially the case as the posterior estimate for ϕ_{kl} is of overall similar size to ϕ_{lc} and ϕ_{kc}. Therefore, I construct a new comparison vector:

$$\theta^* = [\theta_1^*, \theta_2^*, \theta_3^*] = [-1.57, \theta_2^*, \theta_3^*].$$

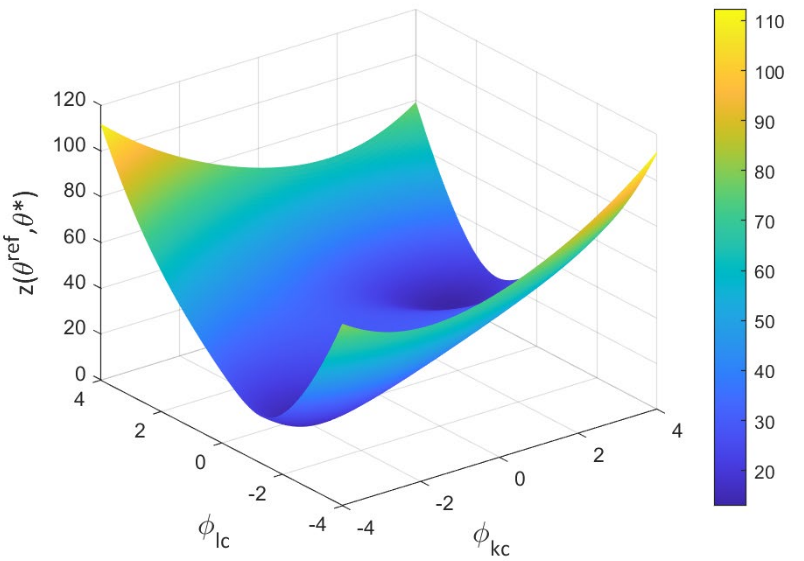
Based on the new θ^*, I repeat the previous analysis comparing the θ^*'s to the same reference mode, $\theta^{ref} = [1.57, -3{,}1]$. The results for the norm of Δ_B and Δ_Ω are presented in Fig. 1.3 and Fig. 1.4.

Fig. 1.3: 3d Graph of normed distance between AR matrices based on alternative mode



Notes: 3d Graph plotting the normed difference, z(θ^{ref}, θ^*), between the autoregressive matrices, B(θ^*) and B(θ^{ref}), over a grid for ϕ_{lc} and ϕ_{kc}. θ^{ref} is set to [1.57, −3,1] and in θ^* the first entry is fixed to -1.57.

Fig. 1.4: 3d Graph of normed distance between Covariance matrices based on alternative mode



Notes: 3d Graph plotting the normed difference, $z(\theta^{ref}, \theta^*)$, between the covariance matrices, $\Omega(\theta^*)$ and $\Omega(\theta^{ref})$, over a grid for ϕ_{lc} and ϕ_{kc}. θ^{ref} is set to $[1.57, -3, 1]$ and in θ^* the first entry is fixed to -1.57.

Firstly, for Δ_B there are four areas that generate matrices similar to the reference mode: $(\phi_{\mathrm{lc}} < 0 \bigwedge \phi_{\mathrm{kc}} < 0)$, $(\phi_{\mathrm{lc}} < 0 \bigwedge \phi_{\mathrm{kc}} > 0)$, $(\phi_{\mathrm{lc}} > 0 \bigwedge \phi_{\mathrm{kc}} < 0)$ and $(\phi_{\mathrm{lc}} > 0 \bigwedge \phi_{\mathrm{kc}} > 0)$. Secondly, for Δ_Ω only the areas of $(\phi_{\mathrm{lc}} < 0 \bigwedge \phi_{\mathrm{kc}} < 0)$ and $(\phi_{\mathrm{lc}} > 0 \bigwedge \phi_{\mathrm{kc}} > 0)$ show similar Covariance matrices. Based on this, there should be two additional modes in the posterior for which the reduced form system is observationally similar. That is at points where both Ω and B are observationally similar, which only holds for $(\phi_{\mathrm{lc}} < 0 \bigwedge \phi_{\mathrm{kc}} < 0)$ and $(\phi_{\mathrm{lc}} > 0 \bigwedge \phi_{\mathrm{kc}} > 0)$. To assess this, Section 1.7 re-estimates the Herbst and Schorfheide (2016) estimation using a significantly increased number of particles and mild prior adjustments. The goal is to ensure that the sampler has sufficient capacity to detect further modes.

1.5 Estimation

For the estimations in this chapter, I reconstruct and extend the data set as described in Leeper, Plante and Traum (2010). The extended data set includes the sample from Q1 1960 up to and including Q4 2018. To allow for direct comparisons to the Leeper, Plante and Traum (2010) and Herbst and Schorfheide (2016) estimations of the model, I restrict the data set to end in 2008Q1 for the replication, multimodality, and main result section. The impact of the data set extension and inclusion of the financial crisis is explored in section 1.9 by examining the stability of the model parameters across time. The data is collected from the Bureau of Economic Analysis (BEA) and the FRED data with some updated data sources and minor changes. See the appendix for a full list of changes and data sources for a detailed description.

Additionally, the appendix shows that all constructed series show a very high correlation with the original Leeper, Plante and Traum data set. Nevertheless, due to data corrections and the adjustments above, differences in the posterior estimates are to be expected. Section 1.6 shows that the posterior estimates based on this data set are close to the estimates in Herbst and

Schorfheide (2016), except for minor differences. For consistency, all further sections use the reconstructed data set.

This chapter features two main estimation techniques: The Random Walk Metropolis-Hastings (RWMH) algorithm and the Sequential Monte Carlo (SMC) Sampler. I use the RWMH-V algorithm for the replication exercise as in Herbst and Schorfheide (2016) and Leeper, Plante and Traum (2010). In general, a RMWH algorithm constructs a chain of draws starting at some initialization and generates a sequence of draws that converges to the posterior distribution under ideal circumstances. The RWMH-V algorithm relies on an initial mode-finding step for the initialization. The posterior is then simulated using the found mode and the Hessian at the mode. RWMH-type algorithms work well if the posterior density is unimodal. If, however, the density is multimodal, RWMH algorithms typically struggle to explore anything but the initial mode. A solution to this problem is the Sequential Monte Carlo (SMC) sampler, as discussed in Herbst and Schorfheide (2016), for the use in DSGE models. The SMC sampler relies on multinomial approximations of a series of densities, typically from the prior to the posterior. If successful, the multinomial approximation can capture multimodality more robustly. Therefore, I apply the SMC sampler in the multimodality section and any further estimations.

Estimation techniques aside, I utilize the highest posterior density intervals and regions as developed in Chen and Shao (1999) and Chen et al. (2000). If the density is multimodal, then a single highest posterior density interval can overstate the uncertainty of the object. This is because it may include all modes and, in addition, the space in between. Separating the interval into two, possibly disjointed regions, as proposed by Chen et al. (2000), resolves this as each region can be tailored to the specific mode.

For the estimations, I use the model files published by Leeper, Plante and Traum (2010). See the appendix for a more detailed description of the estimation techniques and the highest posterior density intervals.

1.6 Replication

This section sets out to address two things. Firstly, I present the results of estimations of the original Leeper, Plante and Traum (2010) model using the original data set and the RWMH-V procedure. Secondly, replication results for the estimation based on the diffuse prior setting with the SMC sampler, as in Herbst and Schorfheide (2016), are discussed.

1.6.1 Prior and estimation detail

The prior distributions for the Leeper, Plante and Traum (2010) are very standard in design and representative of the empirical DSGE literature. Priors for core models are chosen based on economically plausible arguments. Parameters that mostly relate to how well the model is going to fit, like the autoregressive parameters and the standard deviations, are calibrated based on the data construction. Overall, Herbst and Schorfheide (2016) argue that the model prior errs on the side of being tight but economically sound. The prior distributions are presented in Table 1.2.

Focusing on the utility-defining parameters, the consumption substitution parameter, γ, has a mean of 1.75. This implies a substitution effect higher than the logarithmic specification a priori. The prior habit persistence of current consumption to previous aggregate consumption has a mean of 0.5 with a large standard deviation. This ensures that the persistence is both unassuming and diffuse enough to converge to a large range of values. The autocorrelation parameters of the AR(1) process all have a prior mean of 0.5 and a standard deviation of 0.2, allowing for a large range of possible autocorrelation behaviours. However, it necessarily enforces stable eigenvalues. The shock parameters are distributed as an inverse-Gamma with a mean of one and a standard deviation of 4.

The priors for the fiscal parameters, like the core model parameters, are chosen in such a way to embed the model with economic intuition. The image brought forward is a government that strictly saves in high debt periods and provides stimulus in economic downturns. The fiscal parameters can be grouped into three sets. Firstly, the parameters that govern the interactions between the fiscal instruments and debt. These parameters all have a gamma prior with a mean of 0.4 and a standard deviation of 0.2. As the gamma distribution is only defined on positive values, this assumption, in combination with the design of the fiscal rules, ensures that the response of fiscal instruments to rising debt levels is contractionary and deficit reducing.

Table 1.2: Leeper, Plante and Traum (2010) prior

Param.	P.d.f	mean	St.dev	Param.	P.d.f	mean	St.dev	Param.	P.d.f	mean	St.dev
γ_g	$\mathcal{G}$	0.40	0.20	γ	$\mathcal{G}$	1.75	0.50	ρ_{τ^l}	$\mathcal{B}$	0.70	0.2
γ_{tk}	$\mathcal{G}$	0.40	0.20	κ	$\mathcal{G}$	2.00	0.20	ρ_{τ^c}	$\mathcal{B}$	0.70	0.2
γ_{tl}	$\mathcal{G}$	0.40	0.20	h	$\mathcal{B}$	0.50	0.20	ρ_z	$\mathcal{B}$	0.70	0.2
γ_z	$\mathcal{G}$	0.40	0.20	s''	$\mathcal{G}$	5.00	0.50	σ_a	$i\mathcal{G}$	1.00	4.00
φ_{tk}	$\mathcal{G}$	1.00	0.30	δ_2	$\mathcal{G}$	0.70	0.50	σ_b	$i\mathcal{G}$	1.00	4.00
φ_{tl}	$\mathcal{G}$	0.50	0.25	ρ_a	$\mathcal{B}$	0.70	0.2	σ_l	$i\mathcal{G}$	1.00	4.00
φ_g	$\mathcal{G}$	0.07	0.05	ρ_b	$\mathcal{B}$	0.70	0.2	σ_i	$i\mathcal{G}$	1.00	4.00
φ_z	$\mathcal{G}$	0.20	0.10	ρ_l	$\mathcal{B}$	0.70	0.2	σ_g	$i\mathcal{G}$	1.00	4.00
ϕ_{kl}	$\mathcal{N}$	0.25	0.10	ρ_i	$\mathcal{B}$	0.70	0.2	σ_{τ^k}	$i\mathcal{G}$	1.00	4.00
ϕ_{kc}	$\mathcal{N}$	0.05	0.10	ρ_g	$\mathcal{B}$	0.70	0.2	σ_{τ^l}	$i\mathcal{G}$	1.00	4.00
ϕ_{lc}	$\mathcal{N}$	0.05	0.10	ρ_{τ^k}	$\mathcal{B}$	0.70	0.2	σ_{τ^c}	$i\mathcal{G}$	1.00	4.00
								σ_z	$i\mathcal{G}$	1.00	4.00

Notes: For the P.d.f., $\mathcal{G}$ corresponds to a Gamma prior, $\mathcal{B}$ is a Beta prior, $i\mathcal{G}$ is an inverse Gamma distribution, and $\mathcal{N}$ is a normal distribution.

Secondly, the output response parameters are constructed in a similar way with a gamma distribution. The implication based on the fiscal response functions is that the fiscal body reduces expenditure and increases taxation in response to economic growth. Lastly, the interaction parameters are distributed as normal distributions. Here, stronger positive interactions are implied by the prior between capital and labour. Comparatively, weaker interactions between capital and labour/consumption are expected a priori.

The diffuse estimation by Herbst and Schorfheide (2016) explores model dynamics under a comparatively smaller influence of economic guidance on prior design. Changes to priors are restricted to fiscal parameters. The first major change is that now fiscal policy is also allowed to respond positively or negatively to economic growth. Unlike Leeper, Plante and Traum (2010), Herbst and Schorfheide (2016) do not assume that the government necessarily provides stimulus in economic downturns but lets the data speak for itself on the matter.

The assumption of countercyclical debt policy is maintained. Though, an unassuming uniform prior is selected to ensure that the posterior estimates are driven by the likelihood of the model. The last major change is that the standard deviation for all parameters is increased significantly. Table 1.3 presents the prior changes for the diffuse prior.

Table 1.3: diffuse prior changes

Param.	P.d.f	mean	St.dev
γ_g	$\mathcal{U}$	2.5	1.44
γ_{tk}	$\mathcal{U}$	2.5	1.44
γ_{tl}	$\mathcal{U}$	2.5	1.44
γ_z	$\mathcal{U}$	2.5	1.44
φ_{tk}	$\mathcal{N}$	1.00	1.00
φ_{tl}	$\mathcal{N}$	0.50	1.00
φ_g	$\mathcal{N}$	0.07	1.00
φ_z	$\mathcal{N}$	0.20	1.00
ϕ_{kl}	$\mathcal{N}$	0.25	1.00
ϕ_{kc}	$\mathcal{N}$	0.05	1.00
ϕ_{lc}	$\mathcal{N}$	0.05	1.00

Notes: For the P.d.f., $\mathcal{G}$ corresponds to a Gamma prior, $\mathcal{B}$ is a Beta prior, $i\mathcal{G}$ is an inverse Gamma distribution, and $\mathcal{N}$ is a normal distribution.

For the replication of the Leeper, Plante and Traum (2010) estimation, I utilize the RWMH-V algorithm. Herbst and Schorfheide (2016) version is estimated using sequential Monte Carlo techniques. For more detail, see the estimation detail section before.

1.6.2 Posterior estimates

The key result of the posterior estimates is that both estimations using the original and diffuse priors are consistent with the estimates found previously with the respective techniques. All posterior estimates are close, and deviations are generally minor. The estimates are mostly contained in the HPD intervals obtained in the original papers. The results for the fiscal parameters are presented in Table 1.4, and for the remaining parameters, the results are described in Table 1.5 below.

The key difference between the two estimations is the choice of the diffuse prior design. For the fiscal interaction parameters, the replication exercise confirms the finding by Herbst and

Schorfheide (2016) that the diffuse prior results in quite different posterior estimates. The choice of the uniform prior for the debt parameters has resulted in small changes in the posterior estimates, with a minor exception of γ_{tl}. The mean estimates for γ_{tl} are on the boundaries of the HPD intervals of the two estimations but still contained in.

The choice of changing the distribution of the output response parameter to a normal distribution has resulted in the parameters for government consumption and transfers changing sign. The implication is that in partial equilibrium, government consumption and transfers' response is now pro-cyclical. This means that if output increases, the fiscal institute raises expenditure across the board.

Table 1.4: Comparison of fiscal parameter estimates of the replication exercise

Param.	Original prior (RWMH-V) mean [5% - 95% HPD int.]	Diffuse prior (SMC) mean [5% - 95% HPD int.]
γ_g	0.16 [0.07,0.27]	0.11 [0.00,0.25]
γ_{tk}	0.36 [0.19,0.57]	0.37 [0.09,0.67]
γ_{tl}	0.12 [0.04,0.21]	0.04 [0.00,0.12]
γ_z	0.33 [0.18,0.48]	0.33 [0.12,0.54]
φ_{tk}	1.72 [1.21,2.24]	2.14 [1.28,2.89]
φ_{tl}	0.28 [0.11,0.52]	0.11 [-0.45,0.69]
φ_g	0.06 [0.01,0.13]	-0.47 [-1.04,0.10]
φ_z	0.17 [0.06,0.32]	-0.11 [-0.68,0.54]
ϕ_{kl}	0.19 [0.14,0.24]	1.50 [1.18,1.97]
ϕ_{kc}	0.03 [-0.03,0.09]	-0.19 [-3.03,2.99]
ϕ_{lc}	-0.02 [-0.07,0.04]	0.18 [-1.34,1.50]

Notes: The table includes estimates and HPD intervals of key fiscal parameters for both the original prior and diffuse prior

Increasing the prior standard deviation has resulted in some parameters diverging significantly further from the prior mean than before. For example, the posterior estimates for the interaction parameters have changed significantly and take larger values in absolute terms. As such, the interaction between capital and labour taxation is now estimated to be 1.5 under the diffuse priors in comparison to 0.19 under the original priors. The same behaviour applies to the other interaction parameters.

An indirect effect of changing the priors for the fiscal parameters is that the standard deviations for fiscal shocks have decreased. The shock standard deviations for tax variables have decreased the most. Implicitly, this tells us that a part of the dynamic generated by the shock processes under the original priors has been internalized by the modelling framework under the diffuse priors. In turn, this ought to reduce the standard deviation.

Table 1.5: Comparison of the remaining parameter estimates of the replication exercise

Param.	Original prior (RWMH-V) mean [5% - 95% HPD int.]	Diffuse prior (SMC) mean [5% - 95% HPD int.]	Param.	Original prior (RWMH-V) mean [5% - 95% HPD int.]	Diffuse prior (SMC) mean [5% - 95% HPD int.]
ρ_{τ^c}	0.93 [0.89,0.97]	0.96 [0.93,1.00]	γ	2.54 [1.84,3.35]	2.39 [1.48,3.22]
ρ_z	0.95 [0.91,0.98]	0.95 [0.90,0.99]	κ	2.15 [1.48,2.95]	2.08 [1.20,3.02]
σ_a	0.63 [0.57,0.68]	0.63 [0.56,0.70]	h	0.54 [0.43,0.65]	0.54 [0.40,0.67]
σ_b	7.18 [6.44,7.98]	6.98 [6.06,7.87]	s''	5.52 [5.10,5.96]	5.50 [4.97,5.95]
σ_l	2.98 [2.33,3.78]	2.81 [1.96,3.69]	δ_2	0.29 [0.18,0.46]	0.29 [0.14,0.51]
σ_i	6.53 [5.70,7.49]	6.27 [5.21,7.28]	ρ_a	0.97 [0.95,0.99]	0.97 [0.94,0.99]
σ_g	3.06 [2.81,3.33]	2.91 [2.58,3.23]	ρ_b	0.66 [0.62,0.70]	0.65 [0.59,0.70]
σ_{τ^k}	4.43 [4.07,4.83]	1.25 [1.00,1.53]	ρ_l	0.98 [0.96,1.00]	0.98 [0.96,1.00]
σ_{τ^l}	2.98 [2.73,3.25]	2.00 [1.59,2.44]	ρ_i	0.54 [0.45,0.64]	0.53 [0.41,0.65]
σ_{τ^c}	4.03 [3.70,4.38]	1.15 [0.93,1.41]	ρ_g	0.96 [0.94,0.99]	0.96 [0.93,0.99]
σ_z	3.37 [3.10,3.67]	3.33 [3.00,3.68]	ρ_{τ^k}	0.93 [0.88,0.97]	0.94 [0.88,0.99]
			ρ_{τ^l}	0.98 [0.95,0.99]	0.93 [0.86,0.99]

Notes: The table provides estimates and HPD intervals for the non-fiscal parameters of the model for the original and diffuse prior.

1.7 Multimodality and data reconstruction

Section 1.4.2 presented an argument for why there may be additional multimodality in the parameter ϕ_{kl} in combination with the pre-existing multimodality in ϕ_{kc} and ϕ_{lc}. Picking up on this, this section aims to explore whether the graphical analysis can be confirmed. To do so, the diffuse estimation is rerun using small adjustments to the SMC algorithm and the priors to ensure sufficient coverage.

1.7.1 Prior and estimation detail

To further explore the set modes, I adjust the priors chosen by Herbst and Schorfheide (2016) and centre the normal priors of the fiscal interaction parameters at 0. For ϕ_{kc} and ϕ_{lc} this implies a minor mean shift from 0.05 to 0. For ϕ_{kl} the mean shift is larger and from 0.25 to 0. The mean shift is included to shift some of the initial particles closer to the hypothesised $\phi_{\mathrm{kl}} < 0$ modes and to ensure sufficient coverage for the initial approximation. Based on initial tests, the adjustment helps to improve the exploration of the proposed modes. Arguably, this is a relatively minor change for two reasons. The shape and uncertainty of the prior are diffuse, as per Herbst and Schorfheide (2016). In this case, the influence on the posterior is comparatively small. That is because the likelihood significantly outweighs the prior in most DSGE models, and, as such, the impact of the change ought to be relatively small. Fiscal interaction parameters aside, all other priors are consistent with Herbst and Schorfheide (2016).

The model is then estimated using the SMC sampler. Here and going forward, I increase the number of particles to 20000. Increasing the number of particles in difficult posteriors that suffer from multimodality or non-informative likelihoods is beneficial in that the approximation accuracy is increasing in the number of particles. While computationally intensive, it can improve posterior exploration. The model is estimated on both the original Leeper, Plante and Traum data set and the reconstructed set. The former is labelled as Estimation I and the latter as Estimation II.

1.7.2 Posterior estimates

The estimation results are presented in Table 1.6 and Table 1.7. The parameter estimates for most fiscal parameters across estimations I and II are formally very consistent with the estimates found by Herbst and Schorfheide (2016) using $N = 6000$ and the results in section 1.6 above. All posterior point estimates for the fiscal parameters in estimations I and II are contained in the HPD intervals found by Herbst and Schorfheide (2016). The multimodal fiscal interaction parameters differ the most. ϕ_{kc} and ϕ_{lc} deviate from the Herbst and Schorfheide (2016) estimates but are still very comfortable in the HPD interval. Based on repeated runs, it seems to be the case that there is some variation across runs. The posterior estimates depend on how well each mode is discovered in each run.

Table 1.6: Posterior estimate comparison for fiscal parameters for estimation I and II

Param.	Estimation I (original data set) mean [5% - 95% HPD int.]	Estimation II (replicated data set) mean [5% - 95% HPD int.]
γ_g	0.10 [0.00,0.25]	0.11 [0.00,0.25]
γ_{tk}	0.38 [0.07,0.68]	0.28 [0.07,0.52]
γ_{tl}	0.04 [0.00,0.13]	0.03 [0.00,0.12]
γ_z	0.33 [0.10,0.55]	0.16 [0.00,0.43]
φ_{tk}	2.15 [1.33,2.95]	1.94 [1.17,2.73]
φ_{tl}	0.13 [−0.47,0.73]	0.12 [−0.41,0.67]
φ_g	−0.46 [−1.07,0.10]	−0.53 [−1.10,0.06]
φ_z	−0.12 [−0.74,0.53]	0.37 [−0.50,1.22]
ϕ_{kl}	1.44 [−1.54, 2.05]	1.32 [1.00,1.70]
ϕ_{kc}	−0.15 [−3.1,3.09]	0.73 [−3.61,4.02]
ϕ_{lc}	0.12 [−1.49,1.81]	−0.07 [−1.03,1.17]

Notes: Posterior estimate and HPD interval comparison for key fiscal policy parameters across estimations I and II. Estimation I is based on the original data set, and II on the replicated set.

Looking at the remaining model parameters, some subtle differences in the posterior estimates can be observed between estimations I and II and the original estimates by Herbst and Schorfheide (2016). For Estimation I, all remaining parameters except for s'' are at least contained in the HPD interval if not very similar in absolute terms to the original estimates. s'' is estimated to be significantly lower at 5.54 in comparison to 6.9. For estimation II, more differences can be found. Identically to estimation I, s'' is also estimated to be significantly lower at 5.63. Furthermore, the shock standard deviations σ_b and σ_z are both estimated to be higher in the replicated dataset. Lastly, δ_2, which governs the depreciation of capital, is estimated to be significantly higher. The remaining parameters are similar to Herbst and Schorfheide's estimates or contained in the HPD intervals.

Table 1.7 Posterior estimate comparison for the remaining parameters for estimation I and II

Param.	Estimation I (original data set) mean [5% - 95% HPD int.]	Estimation II (replicated data set) mean [5% - 95% HPD int.]	Param.	Estimation I (original data set) mean [5% - 95% HPD int.]	Estimation II (replicated data set) mean [5% - 95% HPD int.]
ρ_{τ^c}	0.96 [0.92,0.99]	0.95 [0.91,0.99]	γ	2.38 [1.51,3.13]	2.11 [1.34,2.90]
ρ_z	0.95 [0.9,0.99]	0.95 [0.91,0.99]	κ	2.01 [1.21,2.77]	2.11 [1.22,2.89]
σ_a	0.63 [0.57,0.70]	0.62 [0.56,0.70]	h	0.53 [0.39,0.67]	0.58 [0.43,0.71]
σ_b	6.96 [5.99,7.82]	9.2 [8.07,10.35]	s''	5.54 [4.95,5.98]	5.63 [5.03,6.03]
σ_l	2.74 [1.99,3.44]	2.79 [1.93,3.58]	δ_2	0.30 [0.15,0.57]	0.76 [0.30,1.50]
σ_i	6.28 [5.26,7.38]	5.24 [4.42,6.27]	ρ_a	0.97 [0.94,1.00]	0.97 [0.95,1.00]
σ_g	2.91 [2.58,3.24]	2.84 [2.54,3.18]	ρ_b	0.65 [0.60,0.71]	0.49 [0.42,0.56]
σ_{τ^k}	1.25 [0.94,1.56]	1.07 [0.85,1.34]	ρ_l	0.98 [0.96,1.00]	0.97 [0.95,1.00]
σ_{τ^l}	1.99 [1.29,2.50]	2.26 [1.83,2.73]	ρ_i	0.53 [0.42,0.66]	0.44 [0.31,0.59]
σ_{τ^c}	1.15 [0.92,1.41]	0.84 [0.66,1.04]	ρ_g	0.96 [0.93,0.99]	0.96 [0.93,0.99]
σ_z	3.34 [2.99,3.71]	4.72 [4.21,5.22]	ρ_{τ^k}	0.94 [0.87,0.99]	0.93 [0.86,0.99]
			ρ_{τ^l}	0.93 [0.86,0.99]	0.95 [0.88,1.00]

Notes: The table offers posterior mean and HPD intervals for the remaining parameters.

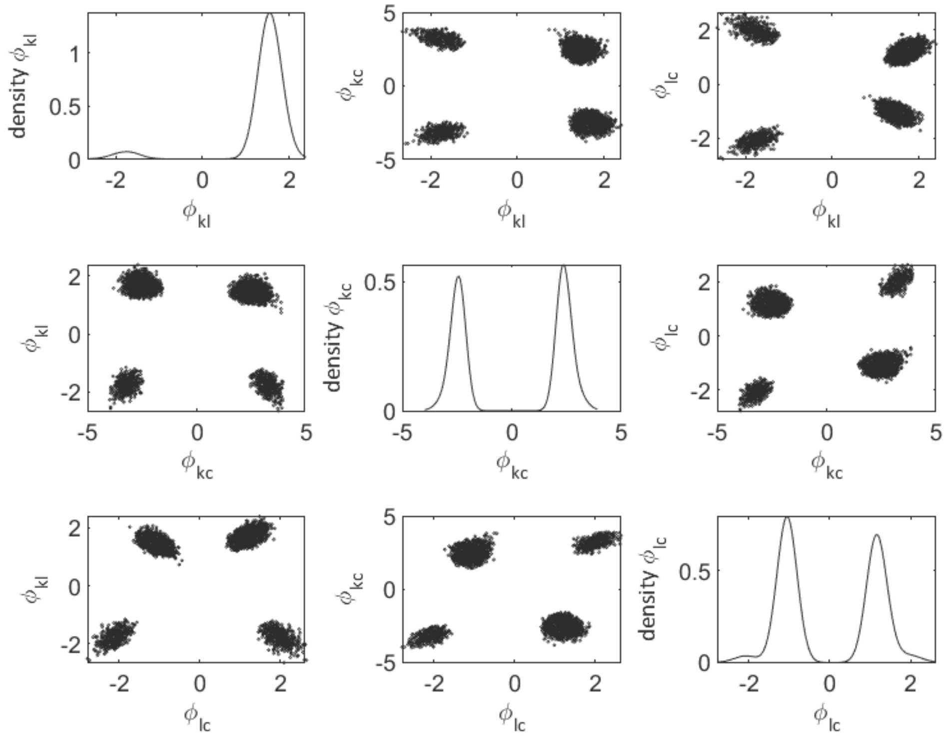
Fig. 1.5 shows posterior density plots and scatter plots of the particle swarm for the fiscal interaction parameters for estimation I. Upon a visual inspection, firstly, the original modes in Herbst and Schorfheide (2016) can be found. Secondly, the existence of the second mode for ϕ_{kl} can be confirmed.

Fig. 1.5 shows a richer picture of multimodal interactions. Originally, the estimation by Herbst and Schorfheide (2016) found modes that were restricted to the following symmetric combinations: $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$ and $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ based on $\phi_{kl} > 0$. This type of mode combination tells a very specific economic story. Fiscal policy at the mode $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$ implies that the consumption tax rate and the labour taxation rate shocks have direct positive impacts on each other. The opposite holds true for consumption taxation and capital taxation. If one analyses the immediate partial equilibrium impact of a positive consumption rate shock on the other taxation rates, one can see that two of the rates increase while one decreases. The key result is that by design, this multimodal relationship is at least partially budget-balancing in spirit. A similar discussion can be made for the mode at $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ with the signs reversed. The results of estimation I show that this behaviour can be extended to a full set of interactions where there is a set of four alternative combinations. To illustrate this, the posterior includes the original modes of $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$ and $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ but also allows for behaviour governed by the remaining two modes: $(\phi_{lc} > 0 \ \wedge \phi_{kc} > 0)$ and $(\phi_{lc} < 0 \ \wedge \phi_{kc} < 0)$ for $\phi_{kl} < 0$. These coincide with the modes proposed in Section 1.4.2 above. What this estimation seemingly confirms is that the multimodality arises based on the design of the fiscal rules, which for the considered range values creates observationally similar laws of motion. There is one caveat, which is that the probability mass for the modes with $\phi_{kl} < 0$ is substantially smaller than for the remaining modes, and the area is less easily explored without prior adjustments. Only around 5.4% of posterior particles sit at the modes with $\phi_{kl} < 0$. Consequently, the additional modes are less relevant in the posterior than the original modes. If the fiscal rules with symmetric interactions are desirable and one wants to have a unimodal posterior, then prior adjustments are the way to obtain this, as the data itself is not sufficient. For example, setting the prior mean for ϕ_{lc} to be larger (smaller) than zero will increase the probability that only the larger (smaller) than zero modes will be explored.

Mechanically, the additional modes define additional dynamics for the DSGE. The mode ($\phi_{lc} > 0 \ \wedge \phi_{kc} > 0$) with $\phi_{kl} < 0$ implies a scenario where, in response to a consumption taxation shock, all other tax rates increase, showing a purely deficit-reducing effect (or expansionary if the shock sign is reversed). While for ($\phi_{lc} < 0 \ \wedge \phi_{kc} < 0$) and $\phi_{kl} < 0$ all interaction parameters are estimated to be negative, and thus, any shock to one rate will force the remaining rates to decrease. The result is that the tax rate interactions in this posterior allow for a more complex set of interactions depending on the particle and perhaps hint at a diverse role of fiscal policy.

The key feature of the increase of particles and prior changes in this section is that it underlines the importance of Sequential Monte Carlo (SMC) techniques for DSGE models. DSGE models offer both the opportunity and difficulty of incredibly expansive modelling and design approaches. The Leeper, Plante and Traum (2015) model is a good example of this with its fiscal interaction parameters. Via multimodality, the fiscal response functions can represent a diverse set of theory-driven behaviours. The difficulty comes into the equation that increasingly complex posterior distributions, in turn, require increasingly elaborate statistical approaches. SMC samplers can fill that role convincingly.

Fig. 1.5 Particle Swarm and density plots for fiscal interaction parameters as in the original model



Notes: For each of the three fiscal interaction parameters, ϕ_{kl}, ϕ_{kc} and ϕ_{lc}, the above Fig. includes one density plot on the diagonal to showcase the multimodality independently of the other parameters. Off-diagonal plots show a scatter plot of the particle swarm for the selected parameters.

1.8 Expanding the fiscal ruleset

The specific assumption made about interactions between fiscal policy tools can be crucial in identifying the dynamics of fiscal policy. Therefore, exploring the dynamics generated by different interaction structures can aid policymakers and economists in their decision processes. After the initial exploration into fiscal rules and the replication, this section sets out to identify the role of alternative fiscal rules with an emphasis on fiscal variable interactions. I construct a set of candidate models to be compared based on a Bayes Factor analysis. For the best fitting models, I explore the effect that the new fiscal rules have on policy interventions and further characterize the role of multimodality using the multimodality HPD regions.

1.8.1 Alternative fiscal rules

Here, I focus on the fiscal rules most commonly found in the literature. To start off with, the exogenous components, $\widehat{Y}_t$ and $\widehat{B}_{t-1}$, to which policymakers respond, are a fixed component across all considered models. This is consistent with the best-fitting model in Leeper, Plante and Traum (2010) and the majority of the surveyed papers in the literature review. Government income variables (i.e. tax rates) and spending variables (transfers and consumption) are viewed as separate blocks. Further, within and across block interactions are considered here. Within the respective budget components, variable interactions are considered to be either independent or to have a symmetric effect on each other (within block).[9] Therefore, four possible fiscal rule models can be constructed from the within assumptions:

Table 1.8: Sample of fiscal rule interactions for within-category interactions

Income rules / spending rules	Independent	Symmetric
independent	$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \hat{B}_{t-1} + \hat{u}_t^{\tau^k}$ $\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \hat{B}_{t-1} + \hat{u}_t^{\tau^l}$ $\hat{\tau}_t^c = \hat{u}_t^{\tau^c}$ $\hat{G}_t = -\varphi_g \hat{Y}_t - \gamma_g \hat{B}_{t-1} + \hat{u}_t^g$ $\hat{Z}_t = -\varphi_z \hat{Y}_t - \gamma_z \hat{B}_{t-1} + \hat{u}_t^z$	$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \hat{B}_{t-1} + \hat{u}_t^{\tau^k}$ $\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \hat{B}_{t-1} + \hat{u}_t^{\tau^l}$ $\hat{\tau}_t^c = \hat{u}_t^{\tau^c}$ $\hat{G}_t = -\varphi_g \hat{Y}_t - \gamma_g \hat{B}_{t-1} + \phi_{gz} \hat{u}_t^z + \hat{u}_t^g$ $\hat{Z}_t = -\varphi_z \hat{Y}_t - \gamma_z \hat{B}_{t-1} + \phi_{gz} \hat{u}_t^g + \hat{u}_t^z$
symmetric	$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \hat{B}_{t-1} + \phi_{kl} \hat{u}_t^{\tau^l} + \phi_{kc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^k}$ $\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \hat{B}_{t-1} + \phi_{lk} \hat{u}_t^{\tau^k} + \phi_{lc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^l}$ $\hat{\tau}_t^c = \phi_{kc} \hat{u}_t^{\tau^k} + \phi_{cl} \hat{u}_t^{\tau^l} + \hat{u}_t^{\tau^c}$ $\hat{G}_t = -\varphi_g \hat{Y}_t - \gamma_g \hat{B}_{t-1} + \hat{u}_t^g$ $\hat{Z}_t = -\varphi_z \hat{Y}_t - \gamma_z \hat{B}_{t-1} + \hat{u}_t^z$	$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \hat{B}_{t-1} + \phi_{kl} \hat{u}_t^{\tau^l} + \phi_{kc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^k}$ $\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \hat{B}_{t-1} + \phi_{lk} \hat{u}_t^{\tau^k} + \phi_{lc} \hat{u}_t^{\tau^c} + \hat{u}_t^{\tau^l}$ $\hat{\tau}_t^c = \phi_{kc} \hat{u}_t^{\tau^k} + \phi_{cl} \hat{u}_t^{\tau^l} + \hat{u}_t^{\tau^c}$ $\hat{G}_t = -\varphi_g \hat{Y}_t - \gamma_g \hat{B}_{t-1} + \phi_{gz} \hat{u}_t^z + \hat{u}_t^g$ $\hat{Z}_t = -\varphi_z \hat{Y}_t - \gamma_z \hat{B}_{t-1} + \phi_{gz} \hat{u}_t^g + \hat{u}_t^z$

Notes: Independent indicates that within a fiscal block (e.g. taxation), all variables are set independently from each other. Symmetric implies the interactions as per the original specification.

The next step in developing a general fiscal interaction ruleset is to look at across interactions between spending and income variables. Here, I focus on block recursive interactions, where the two blocks are the spending and income variables. The idea is to think of the government conceptually choosing, for example, its spending levels first. Based on that, it decides how to

[9] In the baseline Leeper, Plante and Traum (2010) model, the government considers spending variables, transfer and consumption, to be independent from each other and the tax income rates. The tax rates are set jointly via symmetric rules.

adjust tax rates to close the budget. Within the block-recursive structure, each first-ordered variable then has its own independent effect on each second-ordered variable. Below, the independent/independent case with spending ordered first in the block-recursive structure serves as an example:

$$\hat{G}_t = -\varphi_g \hat{Y}_t - \gamma_g \widehat{B}_{t-1} + \hat{u}_t^g$$

$$\hat{Z}_t = -\varphi_z \hat{Y}_t - \gamma_z \widehat{B}_{t-1} + \hat{u}_t^z$$

$$\hat{\tau}_t^k = \varphi_k \hat{Y}_t + \gamma_k \widehat{B}_{t-1} + \phi_{kg} \hat{u}_t^g + \phi_{kz} \hat{u}_t^z + \hat{u}_t^{\tau^k}$$

$$\hat{\tau}_t^l = \varphi_l \hat{Y}_t + \gamma_l \widehat{B}_{t-1} + \phi_{lg} \hat{u}_t^g + \phi_{lz} \hat{u}_t^z + \hat{u}_t^{\tau^l}$$

$$\hat{\tau}_t^c = \phi_{cg} \hat{u}_t^g + \phi_{cz} \hat{u}_t^z + \hat{u}_t^{\tau^c}$$

This type of structure then generates an extra six parameters independently of the ordering. As it is unclear if spending should be ordered first in the structure or taxation, both are considered here. Putting things together, all possible combinations of the afore-described structures are considered. This gives rise to the following set of 12 models:

Table 1.9: Complete description of alternative models

Block recursive	Intra tax	Intra spend.	Number of parameters
Ind.	Ind.	Ind.	31
Ind.	Ind.	Sym.	32
Ind.	Sym.	Ind.	34
Ind.	Sym.	Sym.	35
tax first	Ind.	Ind.	40
tax first	Ind.	Sym.	41
tax first	Sym.	Ind.	43
tax first	Sym.	Sym.	44
spend. first	Ind.	Ind.	40
spend. first	Ind.	Sym.	41
spend. first	Sym.	Ind.	43
spend. first	Sym.	Sym.	44

It is important to note that these results should not be treated as normative statements since the ranking among the fiscal rules may depend on the other components of the model as well, which in turn might be misspecified as well. This is true for any model where we do not know the data generating process. However, given how commonly used a fiscal model like in Leeper, Plante and Traum (2010) is in the literature, it is beneficial to gain a deeper understanding of which fiscal rules fit the data better or what effects the fiscal stimulus has on the economy within that framework.

1.8.2 Prior and estimation detail

The prior construction for the additional parameters is purposefully unassuming and diffuse. For all the additional parameters across the different fiscal rule specifications, the priors are assumed to be normally distributed with mean zero and variance one. Therefore, no assumptions are made about the possible direction of effects. Additionally, for a percentage response parameter, a standard deviation of one covers a large range of possible effects and, as such, can be definitely understood as diffuse.

The models are estimated on the replicated data set on the original time frame from 1960Q1 to 2008Q1. All other estimation details and tuning parameters are as before.

1.8.3 Posterior estimates

This section presents the posterior estimates for the key fiscal interaction structures in the general model space described in the previous section and conducts the Bayes Factor analysis. To focus the analysis on the most important structures, the models are selected based on a posterior odds comparison:

$$\frac{\Pr\ (M_2|D)}{\Pr\ (M_1|D)} = \frac{\Pr(D|M_2) \times \Pr\ (M_2)}{\Pr(D|M_1) \times \Pr\ (M_1)}$$

All models are assumed to be equally likely a priori. Then the posterior odds comparison reduces to the likelihood ratio or Bayes factor, K:

$$K = \frac{\Pr(D|M_2)}{\Pr(D|M_1)}$$

The intuition behind the Bayes factor is based on a relative comparison of the observed data density given the model structures. If the ratio is larger than one, then the density generated by

M_2 is larger than the outcome generated by M_1. In turn, this implies that M_2 is the relatively more credible model. For the baseline model M_1 there are two comparable options. The first option is to choose the most restrictive model with purely independent rules. This model is a natural baseline case and a solid approach to analysing whether easing the restrictions is advantageous. The second option, which is applied here, is to view the alternative specifications relative to the original specification with the symmetric tax rules. The clear advantage of the second option is that it already allows for interactions. Therefore, it increases the burden of evidence required for any given specification. Thus, it also shrinks the number of competitive models. Typically, as in Kass and Raftery (1995), the interpretation of the Bayes factors is phrased in terms of a hypothesis test. If K is sufficiently large, then there is substantial evidence for model two. Here, it is used to also identify comparable models. As such, ratios close to or larger than one are considered for the analysis later in this section.

Table 1.10 presents the log data density, log $(p(Y))$, the difference in log data density to the Leeper, Plante and Traum (2010) model, Δ, and Bayes ratio, K, for all the models considered in the previous section with the original Leeper, Plante and Traum (2010) structure as the M_1 choice.

Table 1.10: Posterior odds comparison

Block recursive	Intra tax	Intra spend.	$\log(p(Y))$	Δ	K
Ind.	Ind.	Ind.	-4124.62	-38.96	0.00
Ind.	Ind.	Sym.	-4127.71	-42.04	0.00
Ind.	Sym.	Ind.	-4085.66	0.00	1.00
Ind.	Sym.	Sym.	-4086.66	-1.00	0.37
tax first	Ind.	Ind.	-4124.01	-38.35	0.00
tax first	Ind.	Sym.	-4126.26	-40.60	0.00
tax first	Sym.	Ind.	-4080.5	5.17	175.32
tax first	Sym.	Sym.	-4084.69	0.97	0.42
spend. first	Ind.	Ind.	-4119.36	-33.69	0.00
spend. first	Ind.	Sym.	-4135.64	-49.98	0.00
spend. first	Sym.	Ind.	-4085.86	-0.19	0.82
spend. first	Sym.	Sym.	-4099.62	-13.96	0.00

Notes: Posterior odds comparison across model categories in Table 10

The first clear result is that the original specification is very competitive in the pool of models considered here. It is easily among the top three models estimated. The top three models within the block structures are (ind., sym., ind.), (tax first, sym., ind.) and (spend. first, sym., ind.). It seems that the combination of symmetric intra-tax rules and independent spending rules is generally preferable. Symmetric intra-tax rules are preferable to their independent counterparts.

The symmetric spending rule structure appears to be disadvantageous and generally generates lower data densities than the independent set. The second result seems to be that the interaction between spending and taxation rulesets can matter strongly. The best-fitting block recursive model with taxes ordered first has very strong evidence in its favour in comparison to all other models. The best-fitting block recursive model with spending ordered first is comparable to the Leeper, Plante and Traum specification with a log data density difference of -0.19.

Moving forward, Table 1.11 presents the posterior estimates for the fiscal parameters for the best fitting models in each of the block recursive categories (independent, tax first, spending first). Model 1 corresponds to the (ind. sym. ind.) as in Leeper, Plante and Traum (2010), Model 2 is the (tax first, sym., ind.) version and Model 3 corresponds to the following ruleset: (spend. first, sym., ind.). The estimates for the original model are as per Section 1.6 above.

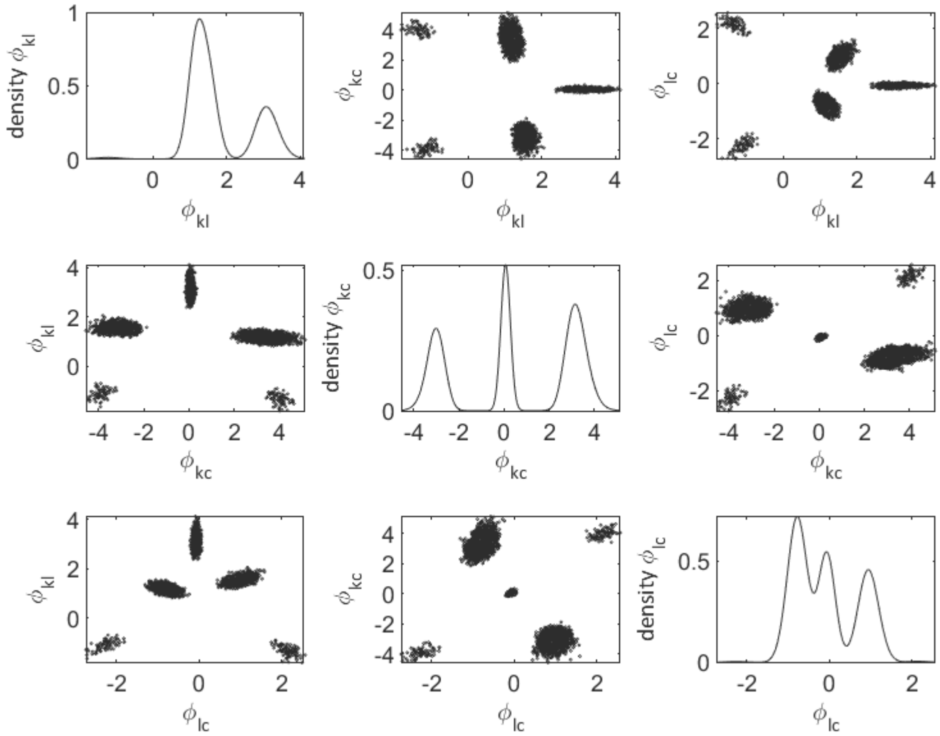
Table 1.11: Posterior estimate comparison for best fitting models by category

Param.	Model 1 (LPT) mean [5% - 95% HPD int.]	Model 2 mean [5% - 95% HPD int.]	Model 3 mean [5% - 95% HPD int.]
γ_g	0.11 [0.00,0.25]	0.11 [0,0.24]	0.11 [0,0.23]
γ_{tk}	0.28 [0.07,0.52]	0.28 [0.08,0.52]	0.29 [0.08,0.51]
γ_{tl}	0.03 [0.00,0.12]	0.03 [0,0.11]	0.04 [0,0.13]
γ_z	0.16 [0.00,0.43]	0.19 [0,0.44]	0.26 [0,0.52]
φ_{tk}	1.94 [1.17,2.73]	2.17 [1.39,2.91]	2.2 [1.43,2.99]
φ_{tl}	0.12 [−0.41,0.67]	0.15 [-0.37,0.71]	0.18 [-0.37,0.73]
φ_g	−0.53 [−1.10,0.06]	-0.46 [-1.05,0.12]	-0.52 [-1.07,0.01]
φ_z	0.37 [−0.50,1.22]	0.55 [-0.29,1.39]	0.44 [-0.36,1.21]
ϕ_{kl}	1.32 [1.00,1.70]	1.35 [1.02,1.76]	1.88 [0.98,3.38]
ϕ_{kc}	0.73 [−3.61,4.02]	-0.35 [-3.84,4.07]	0.47 [-3.54,3.94]
ϕ_{lc}	−0.07 [−1.03,1.17]	0.25 [-0.99,1.24]	-0.07 [-1.01,1.16]
ϕ_{gk}	---	-0.08 [-0.61,0.44]	---
ϕ_{zk}	---	-0.54 [-1.69,0.49]	---
ϕ_{gl}	---	-0.15 [-0.38,0.08]	---
ϕ_{zl}	---	-0.53 [-0.95,-0.1]	---
ϕ_{gc}	---	0.14 [-1.03,1.18]	---
ϕ_{zc}	---	-0.28 [-1.31,0.74]	---
ϕ_{kg}	---	---	-0.36 [-0.58,-0.14]
ϕ_{kz}	---	---	-0.16 [-0.29,-0.04]
ϕ_{lg}	---	---	-0.11 [-0.26,0.04]
ϕ_{lz}	---	---	-0.15 [-0.24,-0.06]
ϕ_{cg}	---	---	0.07 [-0.12,0.27]
ϕ_{cz}	---	---	-0.14 [-0.26,-0.01]

For the original fiscal parameters, the estimates are generally very similar and are easily contained in each other's HPD intervals. The one notable exception to this is ϕ_{kl} under Model 3. The estimate is quite a bit larger, suggesting much stronger interactions between capital and labour tax rates in this specification. Looking at the parameters in Model 2 that make up the block-recursive structure, one can see that the interaction generally appears to be expansionary in the first instance. The point estimates show that in response to a negative structural shock to tax rates, government spending variables typically increase in the first round. One exception is the point estimate for ϕ_{gc}. The same mechanism seems to hold true for Model 2, where the spending variables are ordered first. The parameters that do not enter the fiscal policy rules (i.e. core economic parameters, and specific autoregressive parameters and shock standard deviations) are estimated consistently across the specifications explored here and deviate at most 0.3 standard deviations from the standard Leeper, Plante and Traum (2010) specification. For completeness I provide estimation results for the non-fiscal parameters in the appendix. The next section will then explore the full general equilibrium mechanism using impulse responses.

Further, under Model 2, the multimodality of the fiscal interaction parameters, ϕ_{kl}, ϕ_{lc} and ϕ_{kc} of the diffuse prior estimation as part of the multimodality section is maintained though varied. In fact, in this estimation, the sampler picked up on three modes per parameter, as can be seen in Fig. 1.6 below. Fig. 1.6 shows scatter plots of the particle swarm and density plots for the three interaction parameters. For ϕ_{lc} and ϕ_{kc} two of the modes discovered are consistent with the previous results. The four consistent modes are around ± 1 for ϕ_{lc} and at ± 3 for ϕ_{kc}. Implicitly, these modes describe a fiscal policy that considers setting taxation rates in a joint framework. Changes in one marginal taxation rate can have strong and particularly variable consequences for other rates. In contrast to previous results, both parameters have a third mode, which is illustrated by a significant probability mass around zero. This part of the parameter space is more reflective of a fiscal institution that considers the respective taxation rates independently, which extends the narrative framework for possible types of tax rate interactions.

Fig. 1.6: Particle Swarm and density plots for fiscal interaction parameters for estimation II



Notes: For each of the three fiscal interaction parameters, ϕ_{kl}, ϕ_{kc} and ϕ_{lc}, in estimation II, the above Fig. includes one density plot on the diagonal to showcase the multimodality independently of the other parameters. Off-diagonal plots show a scatter plot of the particle swarm for the selected parameters.

Visually, ϕ_{kl} hardly maintains the multimodality of the previous section, as shown by the small number of particles at the negative mode. The problem here might be that the posterior estimates of the SMC sampler seem to vary mildly across runs for difficult posteriors, and a repeated estimation might offer runs where the negative mode is more fully explored, as in the previous section. In comparison to ϕ_{lc} and ϕ_{kc}, ϕ_{kl} does not have a third mode at zero but at around three. This third mode is expressed in particles in combination with ϕ_{lc} and ϕ_{kc} close to zero. The consequence is that, for these particles, the taxation rates can be separated into labour and capital on the one side and consumption taxation on the other side. Capital and labour taxation rates interact very strongly, as symbolized by the mode of ϕ_{kl} at around three. Opposingly, the consumption taxation rate operates mostly independently from the other two rates.

1.8.4 Impulse responses

In this section, I construct impulse responses for the three best-fitting models. I focus on the effect of fiscal variables on output as a measure of economic performance and the debt level to capture the budgetary impact of each fiscal shock. To condense this room for exploration, I focus on two things. Firstly, I explore the effects that shocks to individual fiscal variables have on output as the key economic measure. Secondly, as opposed to focusing on government variables, here I focus on viewing aggregate budgetary impacts. This reduces the complexity of the interaction to what crucially matters for the policymaker: How much tax income am I going to have, how much is the government going to spend, and what happens to the debt level?

Starting with the government budget constraint, the components can be divided into the aforementioned groups. Firstly, expenditure is defined as $G_t + Z_t$, summing up government consumption and transfers. Secondly, I aggregate capital, labour and consumption tax income to represent the government's total income. The last component is then set to be the debt level B_t. Alternatively, one could choose the government deficit, which is defined as $B_t - B_{t-1}R_{t-1}$. The advantage of the deficit characterization is that it accounts for the interest effects in the refinancing of previous period debt, $B_{t-1}R_{t-1}$. In practice, working with B_t is preferable as it turns out to be more numerically stable. The three objects of interest are:

$$exp_t = G_t + Z_t,$$

$$T_t = T_t^k + T_t^l + T_t^c,$$

$$debt_t = B_t.$$

In log-linear terms, the above budget components can be approximated as follows:

$$exp_t \approx G\left(1 + \hat{G}_t\right) + \mathrm{Z}\left(1 + \hat{Z}_t\right),$$

$$T_t \approx T^k\left(1 + \hat{T}_t^k\right) + T^l\left(1 + \hat{T}_t^l\right) + T^c\left(1 + \hat{T}_t^c\right),$$

$$debt_t \approx B\left(1 + \widehat{B}_t\right).$$

In the final step, the components are viewed in terms of deviations from their steady state:

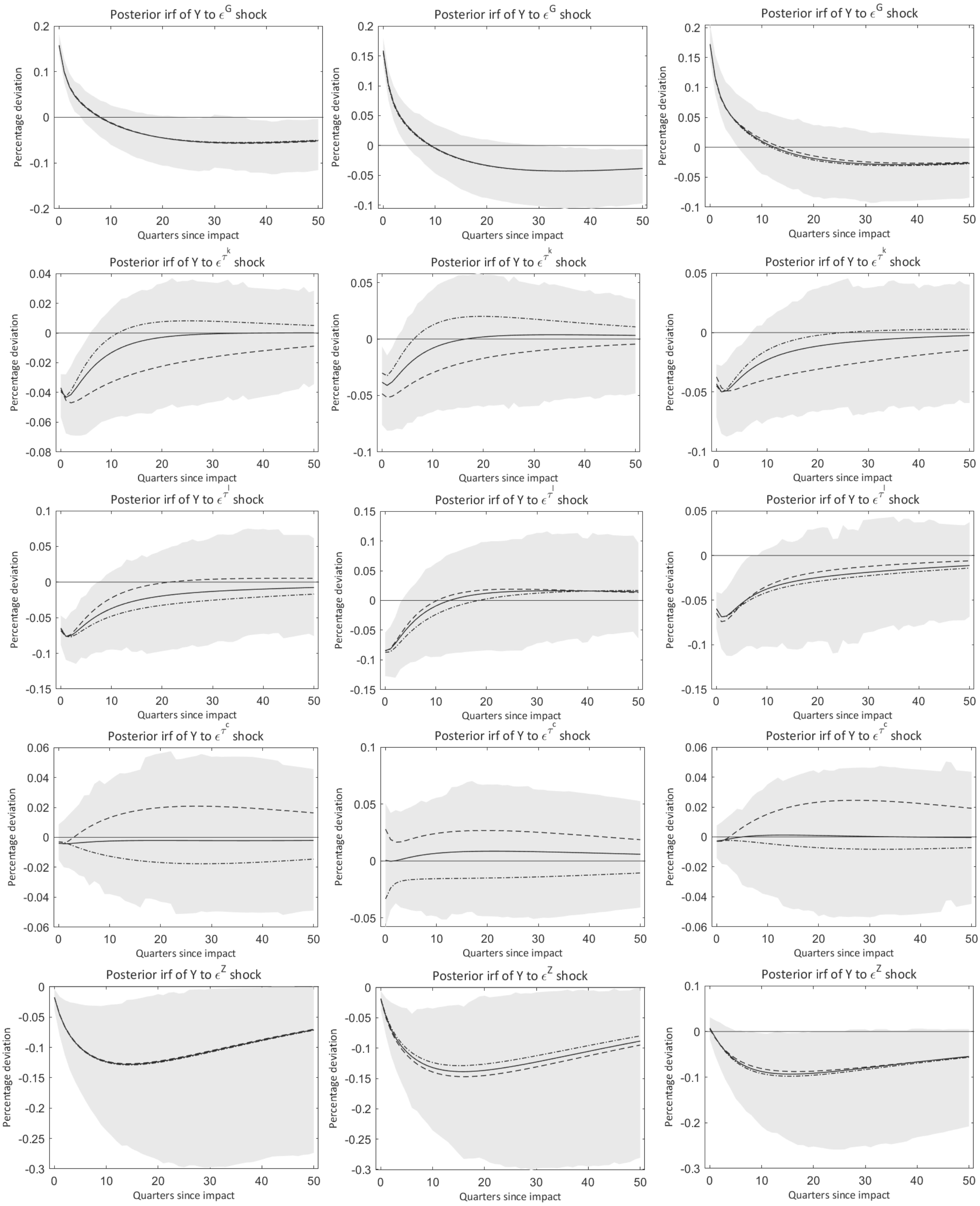
$$\widehat{exp}_t \approx \frac{G}{G+Z}\hat{G}_t + \frac{Z}{G+Z}\hat{Z}_t,$$

$$\hat{T}_t \approx \frac{T^k}{T}\hat{T}_t^k + \frac{T^l}{T}\hat{T}_t^l + \frac{T^c}{T}\hat{T}_t^c,$$

$$\widehat{debt}_t \approx \widehat{B}_t.$$

Fig. 1.7 presents the impulse responses of output in response to fiscal shocks for the three selected models: the original model, taxation first and the spending first variations. The models are coded as (ind., sym., ind.), (taxation first, sym., ind.) and (spending first, sym., ind.), respectively. Fig. 1.8, Fig. 1.9, and Fig. 1.10 present the impulse responses of fiscal budget components to fiscal shocks for the original model, taxation first and the spending first variations, respectively. For all impulse responses, the shock is set to one standard deviation. The time frame is fifty quarters for the impulse responses. Building on previous multimodality results, the resulting consequences for impulse responses are at the centre of the analysis conducted here.

Fig. 1.7: Nominal impulse responses of output to fiscal shocks for selected models



Notes: The figure above displays nominal impulse responses of output to fiscal shocks for the original specification (ind., sym., ind.), (taxation first, sym., ind.) and (spending first, sym., ind.), respectively. The dash-dotted line represents the conditional mean associated with ($\phi_{lc} > 0\ \wedge \phi_{kc} < 0$), and the dashed line is based on particles in ($\phi_{lc} < 0\ \wedge \phi_{kc} > 0$).

To characterize the behaviour of the multimodality, I utilize two approaches. Following Herbst and Schorfheide (2016), the first approach consists of constructing impulse response at different parts of the posterior space of the interaction parameters and analysing the consequences. To do so, I consider the same probability regions as Herbst and Schorfheide (2016) ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$) and ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$), and construct the posterior mean impulse response conditioned on the region. In Fig. 1.7, Fig. 1.8, Fig. 1.9, and Fig. 1.10, the dash-dotted line represents the conditional mean associated with ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$), and dashed line is based on particles in ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$). The advantage of this approach, as this section illustrates, is that impulse responses can behave fundamentally differently based on the particle samples under multimodality.

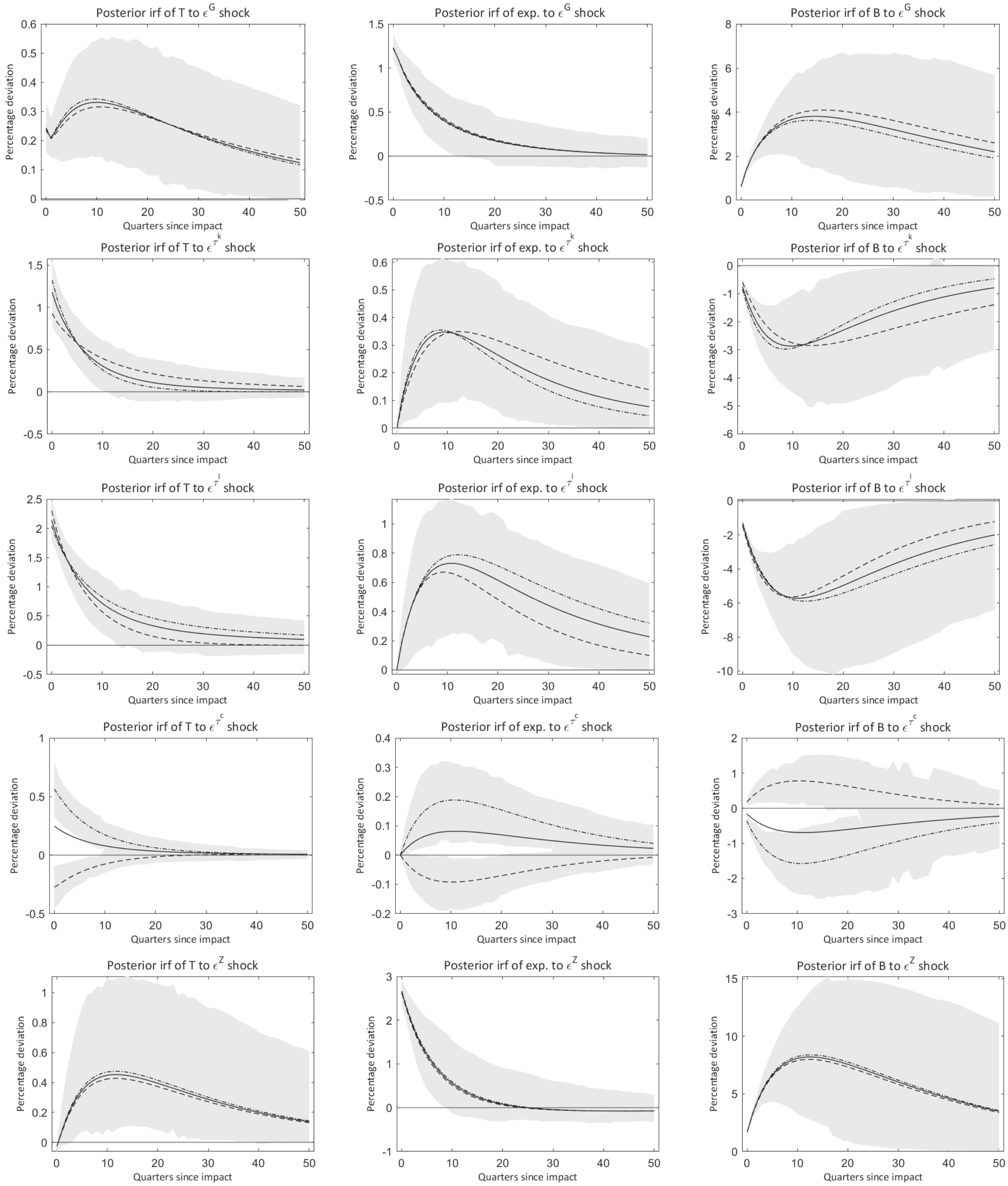
The second approach employed is the construction of the highest posterior density intervals or regions where appropriate. Based on the multimodality of the interaction parameters, it's very intuitive that the impulse responses may also be multimodal, and as such, calculating HPD regions may allow the statistician to correctly assess uncertainty in these cases. In practice, the HPD regions and the condition mean impulse responses show consistent results and complement each other well. In the graphs below, the grey-shaded area corresponds to the 95% HPD interval or regions.

Starting with the analysis, the multimodality of the fiscal interaction parameters can cause multimodal impulse responses in all models in which the conditional mean impulse responses divert significantly from the unconditional mean response. The divergence is not identical for all shocks but depends on the specific fiscal shock. The effect of multimodality is felt strongest in taxation shocks and, specifically, for consumption taxation shocks for all impulse responses. Fiscal expenditure variables, i.e. consumption and transfers, are affected significantly less and typically only divert in the medium to long run.

This creates several unusual analytics about impulse responses. Here, I focus on consumption taxation shocks as they are affected the most by multimodal analytics. To start off, I look at how effective consumption taxation is at stimulating output in Fig. 1.7. A one per cent increase in consumption tax rates can mean very different things for the variables considered here. For the (spending first, sym. and ind.) and the (ind., sym. and ind.) models, the overall mean and the conditional mean impulse responses on output on impact are relatively close to zero. The mean impulse response continues to stay close to zero. Alternatively, the conditional mean impulse responses divert from zero in the medium run. The conditional mean impulse responses based on $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ becomes positive after roughly five quarters, peaking at around a 0.02 steady deviation of output. The other mean, based on $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$, behaves roughly the opposite way, decreasing below zero. Comparing the conditional mean and overall mean impulse responses for this shock, the natural result is that while the unconditional mean implies an inability of consumption tax rates shocks to stimulate the economy, the conditional means tell a story of a persistent and reasonably effective policy tool. The downside seems to be the time until the impulse responses divert from zero. The behaviour of the impulses response for the preferred (taxation first, sym. and ind.) model appears fundamentally different. In comparison to the other two models, the unconditional means deliver non-zero immediate impacts of significant magnitude close to the maximum impacts. After a short dip, they recover to close their maximum impact and then decay slowly, showing significant impacts after 50 quarters.

At first glance, the fact that a positive consumption taxation shock can increase output in a model is counterintuitive. However, for all of the models above, fiscal policy has to be understood as a joint mechanism. Taxation and government expenditures are set jointly. Unless you impose any restrictions on the policy blends, unusual outcomes can be generated. For example, a consumption tax increase may be associated with decreases in the remaining tax rates or other interactions with the expenditure side. In this case, it is not obvious what the effects of the policy blend will be. To ascertain what is happening mechanically, I now look at the budget side to group the effects of the policy blends.

Fig. 1.8: Nominal impulse responses of fiscal budget components to fiscal shocks for the original specification



Notes: The figure above displays nominal impulse responses of fiscal budget components to fiscal shocks for the original specification (ind., sym., ind.). The dash-dotted line represents the conditional mean associated with ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$), and the dashed line is based on particles in ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$).

For the impulse responses of taxation income, the analytics are fairly consistent across models in Fig. 1.8, Fig. 1.9, and Fig. 1.10. On impact, the shock can deliver a significantly positive impact at a roughly 0.50% increase based on the conditional mean of ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$). The behaviour of raising taxation income is consistent with the other tax rates. At the other mode, the impact switches sign and now decreases the tax income to a slightly lesser degree in absolute terms. In terms of the relative weighting of these two scenarios, the posterior mean impact is mildly positive, suggesting that a tax income raising effect on average is more likely.

For the taxation impulse responses, the HPD regions initially separate into two disjointed intervals, and each interval contains one of the conditional mean responses roughly centred in the middle. The separate intervals are relatively tight, and neither contains the unconditional mean nor zero. Therefore, while the unconditional mean measures the average outcome, it is not actually itself a particularly likely outcome. Secondly, the fact that the impulse response HPD regions initially do not include zero implies that the impact is significantly different from zero. The key point is that the model predicts either positive or negative impacts on tax income but not zero impacts.

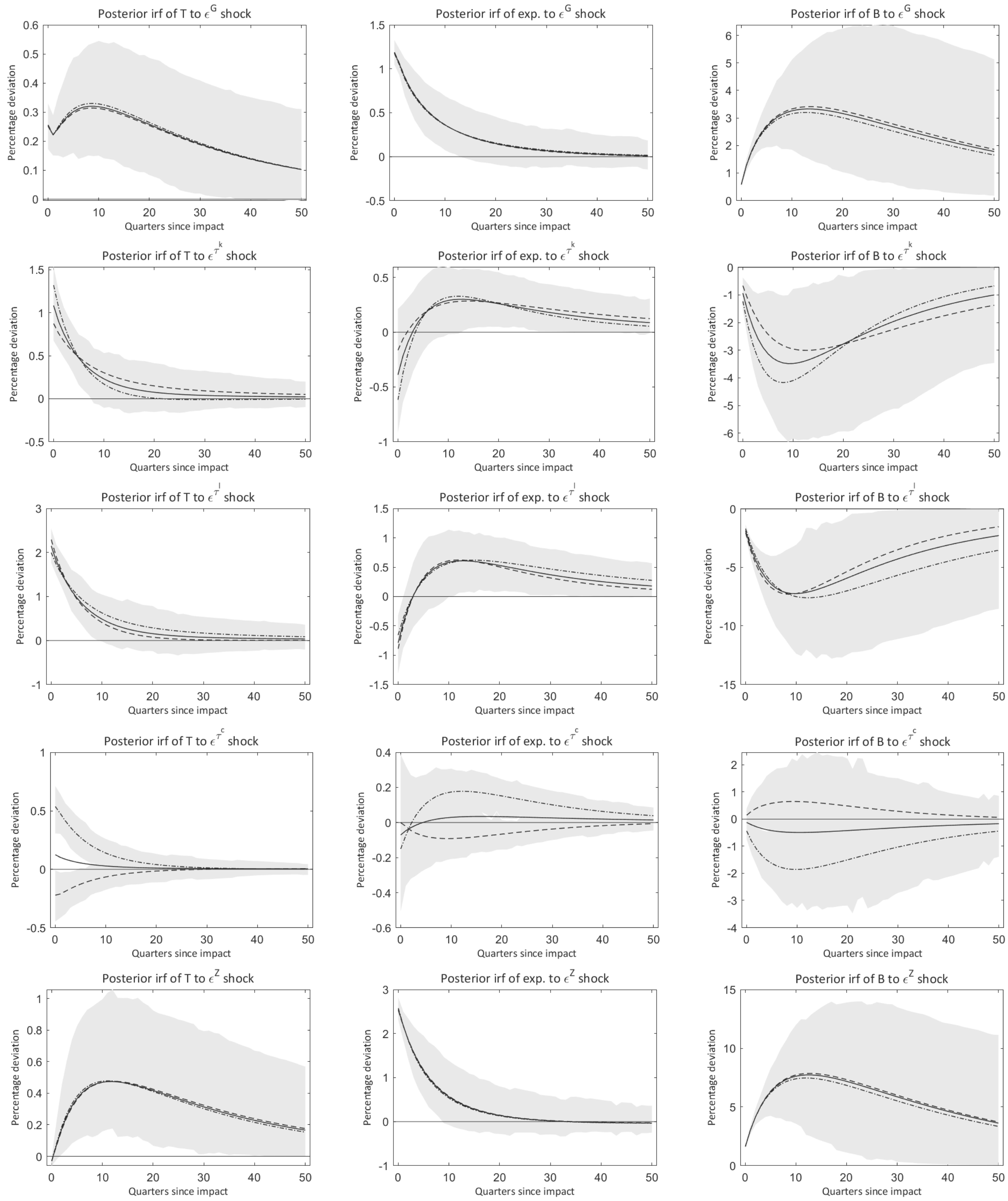
This multimodality trickles down to expenditure as well. On impact, the (spending first, sym. and ind.) and the (ind., sym. and ind.) models predict that the impulse response of expenditure is close to zero for both conditional and unconditional responses. Over time, the conditional responses diverge and peak at ten quarters at 0.2 and -0.1 in comparison to the mean response at 0.8. Between quarters 5 and 30, the HPD regions separate quite strongly, each containing one of the two conditional means. Temporarily, zero is not contained in either HPD region, but this does not last. For the preferred (taxation first, sym. and ind.) model, the situation is mildly different. On impact, the conditional mean based on ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$) delivers a slightly less negative initial impact than its counterpart ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$) or the unconditional mean. The impact based on ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$) is slightly below the unconditional mean. In quarter four, the impulse responses cross over the mean and change their relative position to the mean. Noticeably, the HPD intervals do not separate quite as much as in the previous two cases. The consequence of this multimodal behaviour is that based on the particle, the economic development based on the fiscal shock can be fundamentally different both in the short and long run. Naturally, because taxation and expenditure behave in a multimodal way, debt development is also multimodal.

The same multimodality of impulse responses also applies to capital taxation rate shocks and labour taxation shocks, though less easily visible as the conditional means are much closer to the unconditional mean, and the posterior HPD regions do not separate as much in comparison to the consumption taxation shocks. Nevertheless, the conditional means diverge and seem to further diverge in the medium run. One particular such case is found in the impulse response of output to a capital taxation shock in the original specification in Fig. 1.7. On impact, the conditional and unconditional means are very tightly packed at -0.04. Over time, they separate. The mean response decays slowly from -0.04 to 0. One conditional mean impulse response mirrors this with a much slower decay. The other unconditional mean decays much quicker, crossing over zero at 12 quarters and then continues to deliver positive impacts.

Summing up, multimodality can create ambiguous model analytics that can tell completely different economic stories while still being consistent with the observed data.

For the remainder of this section, I move from the discussion of the consequences of multimodality towards the differences between the models. The main point for the impulse responses on output is that the fiscal interaction structure matters. A one standard deviation shock to government spending delivers a significant 0.15% increase to Y on impact in all models considered and then quickly dissipates. After ten quarters, the mean impact crosses below zero and depresses output, returning to the steady state only very slowly. A one per cent increase in the capital (labour) taxation rate decreases output on impact by 0.04% (0.06%) relative to the steady state, peaking shortly after and then returning to the steady state for all models. Federal consumption taxation rates show a mean response close to zero and never deviate all too far from the steady state. For the conditional means, this changes as previously discussed. The last policy tool to consider is transfers. There seems to be some variation in the impact multipliers to unit transfer shocks. The peak impact for the original parameterisation and the taxation first model is at -0.15% relative to the steady state and at -0.10% for the spending first alternative. The initial impact is similar across all models. Initially, these overall very consistent output multipliers seem puzzling. A reasonable explanation might be that the interaction structure between fiscal instruments matters less for interactions with the economy but could matter for intragovernmental finance. This indeed makes up the next result.

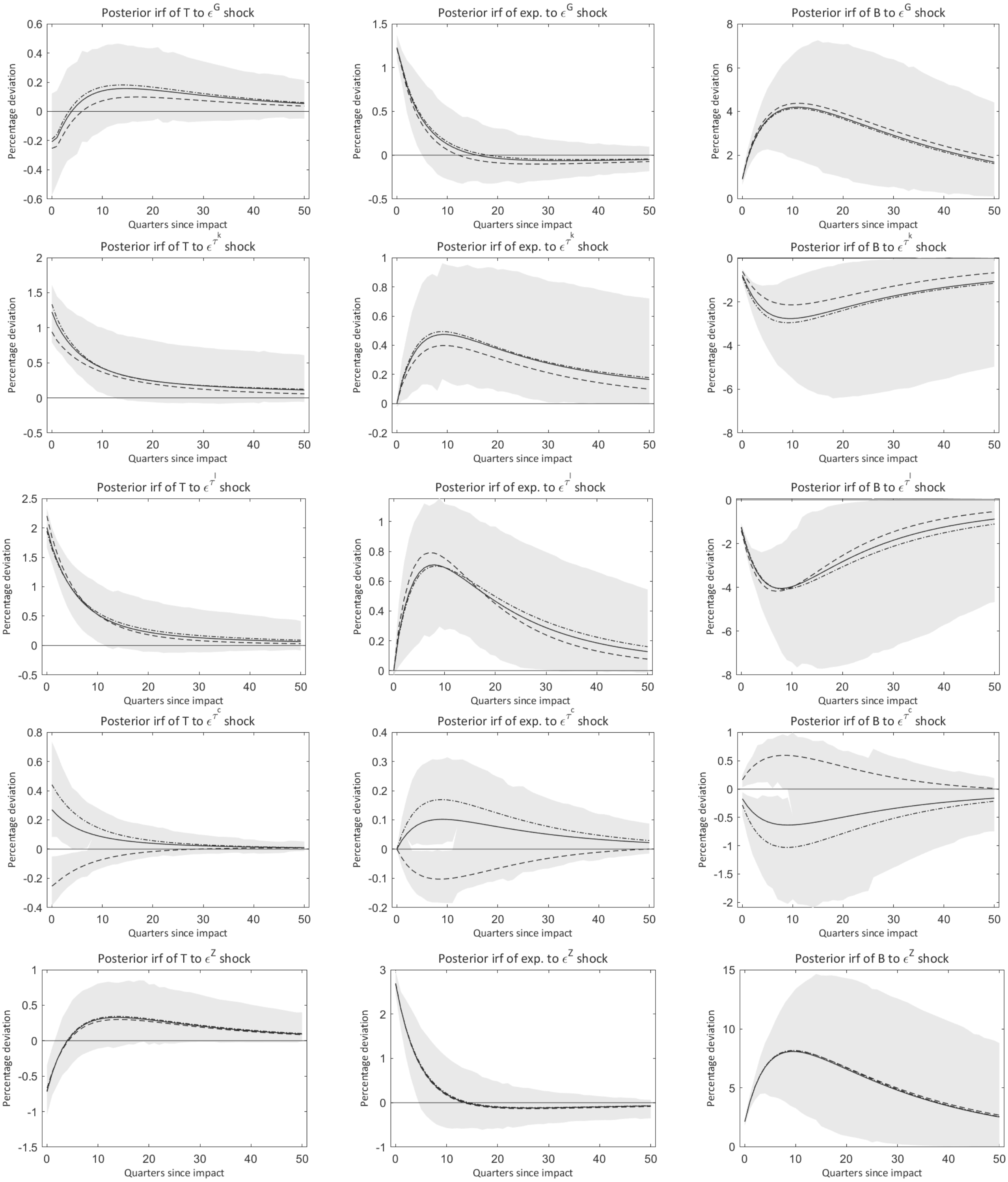
Fig. 1.9: Nominal impulse responses to fiscal shocks for taxation ordered first specification



Notes: The figure above displays nominal impulse responses of fiscal budget components to fiscal shocks for taxation ordered first specification (taxation first, sym., ind.). The line dashed-dotted line represents the conditional mean associated with ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$), and the dashed line is based on particles in ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$).

Budget component financing can change significantly depending on the interaction structure. For the original model, expenditure shocks (spending or transfers) cause a strong and persistent mean increase in the tax income of the government. This suggests that the fiscal institute prefers a selfbalancing expenditure stimulus in comparison to a purely expansionary stimulus. Similarly, in response to taxation shocks, which generally raise the tax income and government expenditure increases. In the two models with extended interaction structures, this changes. Starting with the block recursive ordering with spending first, then tax rules are directly informed by government expenditure shocks within the same period. The result is that in this model, mean taxation income decreases on impact and is below the steady state for roughly eight quarters before switching sign and becoming relatively more budget balancing. The difference between the two parameterisation structures is that under the spending first structure, fiscal stimulus is initially purely expansionary and only becomes balancing later on. A similar dynamic holds for the taxation first model. In the original model, unit shocks to taxation rates cause a very low impact on expenditure on delivery. This then increases a strong positive impact on expenditure. In the taxation first model, government spending rules receive immediate information about taxation shocks. This causes the impulse response of government expenditure to show a negative effect on impact, which only becomes positive later. Therefore, initially, the stimulus becomes purely expansionary from a budget point of view. For consumption taxation shocks, these dynamics are significantly more muddled. The unconditional mean and the conditional mean based on ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$) are consistent with the behaviour described before. The remaining unconditional mean shows persistent deviations below the steady state for both taxation and expenditure. Therefore, while the directions are reverted ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$) shows a budget-balancing behaviour.

Fig. 1.10: Nominal impulse responses to fiscal shocks for spending ordered first specification



Notes: The figure above displays nominal impulse responses of fiscal budget components to fiscal shocks for spending ordered first specification (spending first, sym., ind.). The dash-dotted line represents the conditional mean associated with ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$), and the dashed line is based on particles in ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$).

1.8.5 Present Value Multipliers

Present value multipliers are useful tools in summarizing the economic impacts of stimulus. The ratio that makes up the multiplier is a complex object based on the model dynamics and is particularly well suited to address the dynamic nature of fiscal policy and deficit financing. Following Mountford and Uhlig (2009), the present-value multipliers for output are defined as:

$$PV(X)_k = \frac{\sum_{i=0}^{k}(\prod_{j=0}^{i} r_{t+j}^{-1}) \bigtriangleup Y_{t+i}}{\sum_{i=0}^{k}(\prod_{j=0}^{i} r_{t+j}^{-1}) \bigtriangleup X_{t+i}}.$$

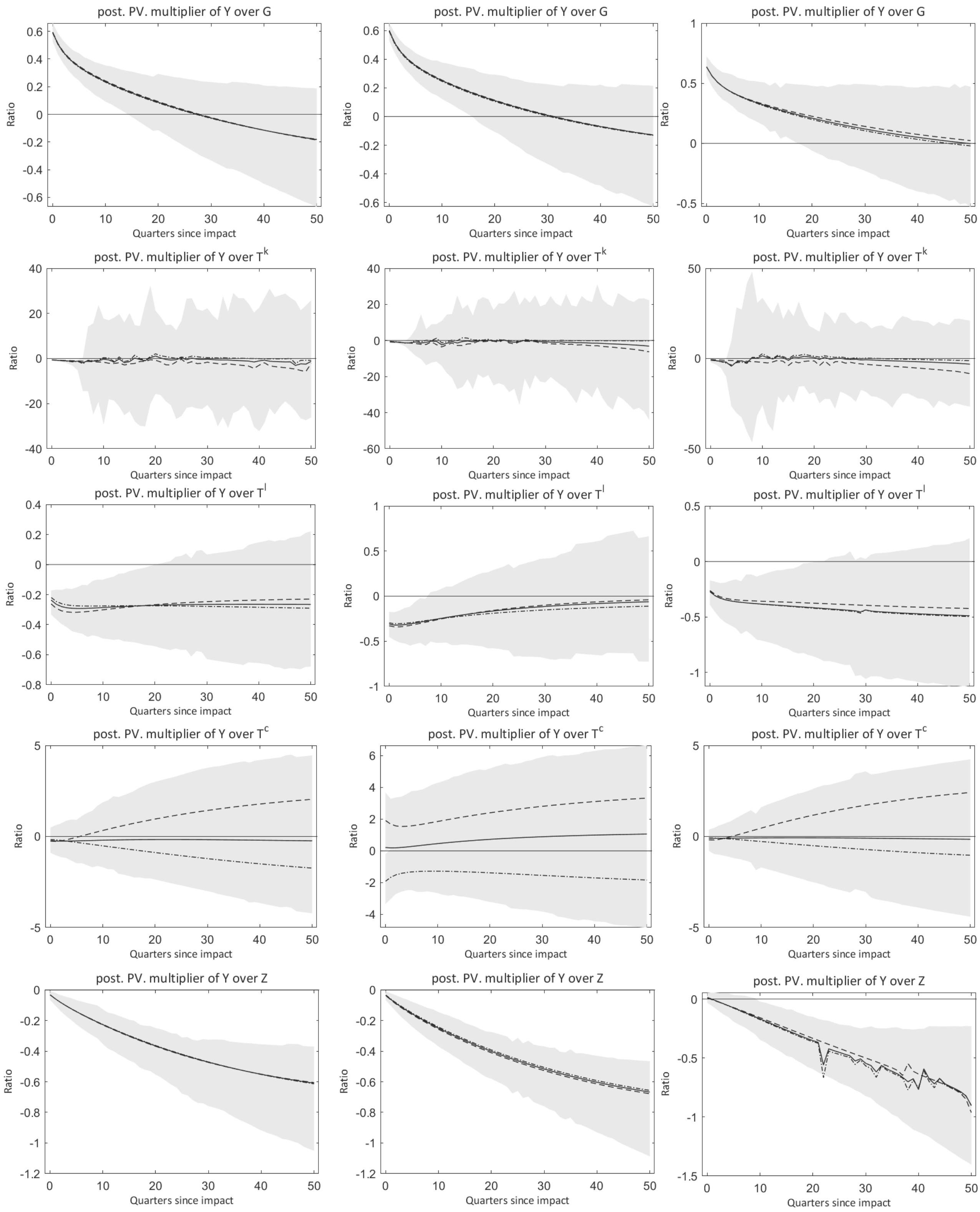
X describes the fiscal variable of interest. $\bigtriangleup$ denotes level changes of the respective variables from their steady state. The discount factor $(\prod_{j=0}^{i} r_{t+j}^{-1})$ is generated based on the dynamically generated interest paths implied by the initial fiscal shock.

Fig. 1.11 shows present value multipliers of output for all five fiscal shock types and their corresponding budget component across the three selected models: the original model, taxation first and the spending first variations. The time frame chosen for the analysis of the multipliers is 50 quarters. The behaviour of the multiplier can change dramatically over its lifetime, and therefore, a long but finite timeframe is considered. As in the previous section, particles at $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$ and $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ are important, and I construct the posterior PV paths conditioned on the region. In Fig. 1.11 below, the dash-dotted line represents the conditional mean associated with $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$, and the dashed line is based on particles in $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$. Additionally, the grey-shaded area corresponds to the 95% HPD interval of the multiplier.

Starting with government consumption multipliers, one can see that the multiplier estimates are fairly consistent with the literature both on impact and in general shape. On impact, all models deliver a multiplier of roughly 0.6, which is consistent with the result of Leeper, Plante and Traum (2010). Over time, the ratio decreases and eventually becomes negative. This behaviour is common to consumption multipliers and has been observed in Zubairy (2014) and Forni et al. (2009). The rate of the decay of the multiplier varies across models. The original model and the taxation first model behave roughly, decaying at similar rates and switching to negative at 35 quarters (8.75 years). The spending first model decays at a much slower rate and only crosses over at 50 quarters. This heavily hints at a different financing scheme in this model.

For labour taxation multipliers, the results are again consistent with Leeper, Plante and Traum (2010) on impact sitting at -0.2. The original specification and the spending first share a common trait for labour taxation multipliers in that after a short period of negative growth over five to ten quarters, the multipliers stabilize close to 0.3 and 0.4, respectively. In contrast, the taxation first multiplier decays back to zero. Since the taxation first model is the preferred model based on the data density, this result paints a less optimal role of labour taxation multiplier than its counterparts. The transfer multipliers are consistent across models and show a very common pattern: on impact, the multiplier is close to zero, and over time, it grows negatively to -0.8 at 50 quarters in.

Fig. 1.11: Real Present Value Multiplier of output to fiscal shocks for selected specifications



Notes: Real Present Value Multiplier of output to fiscal shocks for the original specification (ind., sym., ind.), (taxation first, sym., ind.) and (spending first, sym., ind.), respectively. The dash-dotted line represents the conditional mean associated with ($\phi_{lc} > 0 \ \wedge \phi_{kc} < 0$), and the dashed line is based on particles in ($\phi_{lc} < 0 \ \wedge \phi_{kc} > 0$).

For the consumption taxation multiplier, it makes sense to revisit the topic of multimodality. These multipliers are affected most strongly by multimodality as before. In fact, for the other fiscal variables, it plays a minor role, and deviations only occur in the long run, if at all. The behaviour of consumption multipliers is very similar to the standard impulse responses in the previous section. On impact, the original model and the spending first model show a conditional and unconditional mean multiplier close to zero. The unconditional mean stays close to zero. In contrast, the conditional multipliers based on $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$ and $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ diverge and peak at large negative and positive values, respectively. In contrast to that, the taxation first model delivers a strong initial multiplier for the conditional means of $(\phi_{lc} > 0 \ \wedge \phi_{kc} < 0)$ and $(\phi_{lc} < 0 \ \wedge \phi_{kc} > 0)$ at -2 and 2, respectively. After a short dip, the multiplier then stabilizes around this value. Building on previous results, the multipliers reflect the multimodal behaviour of the impulse response and, therefore, illustrate the importance of this type of analysis.

The odd one out in this analysis is the capital taxation multipliers. They appear relatively unstable and do not show the smoothness common to the other multipliers. While visually unappealing, the reason for this behaviour is simple. The discounted sum of steady state deviations of capital taxation changes the sign at different points over the horizon:

$$\sum_{i=0}^{k}(\prod_{j=0}^{i} r_{t+j}^{-1}) \triangle {T^k}_{t+i}.$$

During the cross-over, the sum in the denominator becomes close to zero. This inflates the multiplier temporarily and explains the abrupt change of scale of the multiplier. This can happen at varying time frames, explaining the repeated undesirable behaviour.

1.9 Sub-sample testing

Fiscal policy can be understood as a slow but changing process of rules and regulations that develops over time. Fiscal policy is typically not time-invariant, and as such, sub-sample estimations can tell us a lot about its development and the variability in estimates seen in the literature. To analyse this, I employ a rolling window subsample estimation procedure to explore the overall development of the parameters that govern the fiscal feedback rules over time for the preferred model (tax first, sym., ind.).

The window size, m, is set to 20 years, and the roll-forward step is set to 10 years. Therefore, two bordering windows share half of each other's observations to ensure some smoothness in the transitions. The initial observation for the first window is set to Q4 1968. The resulting subsamples are Q4 1968 – Q4 1988, Q4 1978 – Q4 1998, Q4 1988 – Q4 2008 and Q4 1998 – Q4 2018.

Table 1.12 Posterior estimate comparison across windows

Param.	Subsample 1 Q4 1968 – Q4 1988 mean [5% - 95% HPD int.]	Subsample 2 Q4 1978 – Q4 1998 mean [5% - 95% HPD int.]	Subsample 3 Q4 1988 – Q4 2008 mean [5% - 95% HPD int.]	Subsample 4 Q4 1998 – Q4 2018 mean [5% - 95% HPD int.]
γ_g	0.07 [0.00,0.19]	0.11 [0.00,0.28]	0.29 [0.00,0.57]	0.32 [0.03,0.62]
γ_{tk}	0.15 [0.00,0.37]	0.43 [0.15,0.76]	0.71 [0.22,1.21]	0.74 [0.30,1.18]
γ_{tl}	0.09 [0.00,0.24]	0.06 [0.00,0.15]	0.25 [0.00,0.53]	0.08 [0.00,0.22]
γ_z	0.28 [0.00,0.57]	0.07 [0.00,0.22]	0.05 [0.00,0.18]	0.19 [0.00,0.53]
φ_{tk}	1.89 [0.85,2.94]	2.15 [1.09,3.21]	2.60 [1.39,3.68]	2.7 [1.42,3.93]
φ_{tl}	-0.24 [-0.98,0.55]	0.00 [-0.49,0.53]	-0.04 [-0.77,0.74]	0.2 [-0.62,1.02]
φ_g	-0.17 [-1.00,0.64]	0.13 [-0.74,0.97]	-0.35 [-1.22,0.59]	-0.26 [-1.31,0.83]
φ_z	0.94 [-0.16,2.10]	0.35 [-0.69,1.40]	1.45 [0.36,2.41]	0.94 [-0.41,2.24]
ϕ_{kl}	0.85 [-1.08,1.61]	-0.15 [-0.53,0.20]	0.99 [-1.36,2.06]	1.65 [1.28,2.03]
ϕ_{kc}	0.15 [-4.39,4.23]	4.45 [3.58,5.33]	0.58 [-3.97,4.21]	1.34 [-3.43,3.74]
ϕ_{lc}	-0.30 [-2.47,2.23]	1.24 [0.71,1.81]	0.34 [-1.00,1.74]	-0.19 [-0.90,0.97]
ϕ_{gk}	0.00 [-1.08,0.95]	0.10 [-0.46,0.70]	-0.31 [-1.20,0.55]	-0.70 [-1.52,0.07]
ϕ_{zk}	-0.48 [-1.91,1.64]	-0.31 [-1.14,0.57]	-0.58 [-1.68,0.72]	-1.36 [-2.74,0.39]
ϕ_{gl}	0.01 [-0.49,0.71]	0.14 [-0.38,0.65]	-0.33 [-0.76,0.08]	-0.50 [-0.84,-0.16]
ϕ_{zl}	-0.67 [-1.40,-0.09]	-0.86 [-1.54,-0.13]	-0.69 [-1.28,0.05]	-1.20 [-1.85,-0.57]
ϕ_{gc}	0.03 [-1.37,1.53]	-1.27 [-2.08,-0.47]	-0.15 [-0.99,0.68]	-0.12 [-0.81,0.56]
ϕ_{zc}	0.16 [-1.94,1.99]	0.27 [-0.72,1.17]	-0.32 [-1.86,1.07]	-0.52 [-1.87,0.97]

Table 1.12 shows posterior estimates for the fiscal feedback rule parameters across the different subsamples. Among the debt-responsive parameters, the more striking results arise. Starting with government consumption, the conclusion can be drawn that the fiscal tool has become significantly more responsive to debt levels. The debt responsiveness parameter γ_g has increased from 0.07 to 0.32 over time. The effect is that government consumption is reduced significantly stronger in response to rising debt levels in later samples. The same applies to the capital taxation rate for which γ_{tk} increased from 0.15 to 0.74. Thus, in later years, the capital taxation rate seems to rise significantly stronger in response to changes in the debt level. The remaining debt parameters vary across the samples but do not offer any trending behaviours.

For the output response parameters, only φ_{tk} shows a sign of a clear trend. Over the subsamples φ_{tk} grows, implying that in later subsamples, the taxation rates' direct response to output growth is to increase stronger and stronger.

For intra-tax interaction parameters, ϕ_{kl}, ϕ_{kc} and ϕ_{lc}, there is not particularly strong trending behaviour to be observed. It is notable that all parameters experience strong variation in their posterior mean estimates across subsamples.

Moving onto the block recursive parameters, ϕ_{gk}, ϕ_{zk}, ϕ_{gl}, ϕ_{zl}, ϕ_{gc} and ϕ_{zc}, they do not show a uniform trend behaviour. Though, what is apparent is that during the two most current subsamples, all parameters have negative posterior mean estimates, while before, the mean estimates were of mixed signs. If the posterior mean estimate is positive, then a unit structural tax shock has a direct positive impact on the respective expenditure variable. The consequence is that the tax shock is partly covered by the government expenditure variable moving to balance the budget. The consequence of negative estimates is that in response to structural tax shocks, the direct interaction causes a decrease in government consumption and transfers at the posterior mean. This type of policy is consistent with purely expansionary fiscal policy based on the interaction of fiscal instruments. The strong shift to negative posterior estimates across the board can tell us that fiscal policy seems to have embraced these purely expansionary interactions across the budget components.

1.10 Conclusion

This chapter has focused on exploring the utility of across-budget block interactions in the fiscal rule set using a data density comparison. The results show that across-block interactions with taxation ordered first can be useful in improving the model fit. Furthermore, across-block interactions in the fiscal rules cause policy interventions to show increased coordination. For example, labour and capital tax rate cuts in the best-fitting model are estimated to be associated with simultaneous but temporary increases in government expenditure. Both measures together imply that the government is conducting purely expansionary stimulus. Only after this initial period does the government start worrying about closing the budget. These results should be interpreted with a grain of salt as this ranking of fiscal rules might change if we use a different DSGE model. Nevertheless, the analysis presented in this chapter provides insights for a widely used fiscal model.

Herbst and Schorfheide (2016) first showed that some parameters in the Leeper, Plante and Traum (2010) model can be estimated as multimodal under diffuse priors and that impulse responses to consumption taxation shocks can consequently also become multimodal. Picking up on this, this chapter showed that the existence of multimodality is a by-product of the design of the original fiscal rules and not, by itself, a feature of the data. Furthermore, using the methodologies in Chen and Shao (1999) and Chen et al. (2000) for evaluating highest posterior density regions, this chapter shows that multimodality can lead to not only different but disjointed effects of policy interventions.

One potential limitation of the fiscal rules proposed in this chapter is that fiscal policy is constrained to act the same regardless of the business cycle conditions. However, it seems reasonable, if not desirable, for fiscal policy to act depending on the circumstances of the economy or the government's financial situation. Chapter 2 approaches this by creating a DSGE model with a fiscal block that is allowed to vary its responses to the business cycle depending on selected economic factors.

1.11 References

Andreyev, M. (2020) Adding a fiscal rule into a DSGE model: How much does it change the forecasts?, Bank of Russia. Available at: https://www.cbr.ru/eng/ec_research/adding-a-fiscal-rule-into-a-dsge-model-how-much-does-it-change-the-forecasts-/ (Accessed: 24 May 2023).

Aursland, T.A. et al. (2020) 'State-dependent fiscal multipliers in Nora - a DSGE model for fiscal policy analysis in Norway', Economic Modelling, 93, pp. 321–353. doi:10.1016/j.econmod.2020.07.017.

Babecký, J., Franta, M. and Ryšánek, J. (2018) 'Fiscal policy within the DSGE-VAR framework', Economic Modelling, 75, pp. 23–37. doi:10.1016/j.econmod.2018.06.005.

Bhattarai, K. and Trzeciakiewicz, D. (2017) 'Macroeconomic impacts of fiscal policy shocks in the UK: A DSGE analysis', Economic Modelling, 61, pp. 321–338. doi:10.1016/j.econmod.2016.10.012.

Bi, H., Shen, W. and Yang, S.-C.S. (2016) 'Fiscal limits in developing countries: A DSGE approach', Journal of Macroeconomics, 49, pp. 119–130. doi:10.1016/j.jmacro.2016.06.002.

Bondzie, E.A. and Armah, M.K. (2022) 'A DSGE model of fiscal stabilizers and informality in Sub-Sahara Africa', Cogent Economics and Finance, 10(1). doi:10.1080/23322039.2022.2137985.

Bureau of Economic Analysis (no date) U.S. Bureau of Economic Analysis (BEA). Available at: https://www.bea.gov/ (Accessed: 17 May 2023).

Bušs, G. and Grüning, P. (2023) 'Fiscal Dsge model for Latvia', Baltic Journal of Economics, 23(1). doi:10.1080/1406099x.2023.2173915.

Cappe, O., Godsill, S.J. and Moulines, E. (2007) 'An overview of existing methods and recent advances in sequential Monte Carlo', Proceedings of the IEEE, 95(5), pp. 899–924. doi:10.1109/jproc.2007.893250.

Carvalho, F.A. and Castro, M.R. (2017) 'Macroprudential policy transmission and interaction with fiscal and monetary policy in an emerging economy: A DSGE model for Brazil', Macroeconomics and Finance in Emerging Market Economies, 10(3), pp. 215–259. doi:10.1080/17520843.2016.1270982.

Cavalcanti, M.A. and Vereda, L. (2015) ‘Fiscal policy multipliers in a DSGE model for Brazil’, Brazilian Review of Econometrics, 35(2), p. 197. doi:10.12660/bre.v35n22015.57570.

Chan, Y.T. (2020) ‘Are macroeconomic policies better in curbing air pollution than environmental policies? A DSGE approach with carbon-dependent fiscal and monetary policies’, Energy Policy, 141, p. 111454. doi:10.1016/j.enpol.2020.111454.

Chen, M.-H. and Shao, Q.-M. (1999) ‘Monte Carlo estimation of bayesian credible and HPD Intervals’, Journal of Computational and Graphical Statistics, 8(1), p. 69. doi:10.2307/1390921.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J.G. (2000) Monte Carlo Methods in bayesian computation. New York: Springer.

Davig, T. and Leeper, E.M. (2011) ‘Monetary–fiscal policy interactions and fiscal stimulus’, European Economic Review, 55(2), pp. 211–227. doi:10.1016/j.euroecorev.2010.04.004.

Drautzburg, T. (2020) ‘A narrative approach to a fiscal DSGE model’, Quantitative Economics, 11(2), pp. 801–837. doi:10.3982/qe1083.

Drautzburg, T. and Uhlig, H. (2015) ‘Fiscal stimulus and distortionary taxation’, Review of Economic Dynamics, 18(4), pp. 894–920. doi:10.1016/j.red.2015.09.003.

Drygalla, A., Holtemöller, O. and Kiesel, K. (2018) ‘The effects of fiscal policy in an estimated DSGE model—the case of the German stimulus packages during the Great Recession’, Macroeconomic Dynamics, 24(6), pp. 1315–1345. doi:10.1017/s1365100518000858.

Engler, P. and Tervala, J. (2018) ‘Hysteresis and fiscal policy’, Journal of Economic Dynamics and Control, 93, pp. 39–53. doi:10.1016/j.jedc.2018.02.002.

Faria-e-Castro, M. (2021) ‘Fiscal policy during a pandemic’, Journal of Economic Dynamics and Control, 125, p. 104088. doi:10.1016/j.jedc.2021.104088.

Federal Reserve Economic Data: Fred: St. louis fed (no date) FRED. Available at: https://fred.stlouisfed.org/ (Accessed: 17 May 2023).

Fernández-Villaverde, J. et al. (2015) ‘Fiscal volatility shocks and economic activity’, American Economic Review, 105(11), pp. 3352–3384. doi:10.1257/aer.20121236.

Forni, L., Monteforte, L. and Sessa, L. (2009) ‘The general equilibrium effects of fiscal policy: Estimates for the Euro Area’, Journal of Public Economics, 93(3–4), pp. 559–585. doi:10.1016/j.jpubeco.2008.09.010.

Fève, P. and Sahuc, J.-G. (2016) ‘In search of the transmission mechanism of fiscal policy in the Euro Area’, Journal of Applied Econometrics, 32(3), pp. 704–718. doi:10.1002/jae.2517.

Gadatsch, N., Hauzenberger, K. and Stähler, N. (2016) ‘Fiscal policy during the crisis: A look on Germany and the euro area with gear’, Economic Modelling, 52, pp. 997–1016. doi:10.1016/j.econmod.2015.10.038.

Gomes, S. et al. (2015) ‘Global policy at the zero lower bound in a large-scale DSGE model’, Journal of International Money and Finance, 50, pp. 134–153. doi:10.1016/j.jimonfin.2014.09.005.

Herbst, E. and Schorfheide, F. (2014) ‘Sequential Monte Carlo sampling for DSGE models’, Journal of Applied Econometrics, 29(7), pp. 1073–1098. doi:10.1002/jae.2397.

Herbst, E.P. and Schorfheide, F. (2016) Bayesian estimation of DSGE models. Princeton: Princeton University Press.

Jesus, D.P., Besarria, C. da and Maia, S.F. (2020) ‘The macroeconomic effects of monetary policy shocks under fiscal constrained’, Journal of Economic Studies, 47(4), pp. 805–825. doi:10.1108/jes-01-2019-0011.

Kang, H. and Suh, H. (2017) ‘Macroeconomic Dynamics in Korea during and after the Global Financial Crisis: A bayesian dsge approach’, International Review of Economics and Finance, 49, pp. 386–421. doi:10.1016/j.iref.2017.02.008.

Kass, R.E. and Raftery, A.E. (1995) ‘Bayes factors’, Journal of the American Statistical Association, 90(430), pp. 773–795. doi:10.1080/01621459.1995.10476572.

Kliem, M., Kriwoluzky, A. and Sarferaz, S. (2016) 'Monetary–fiscal policy interaction and fiscal inflation: A tale of three countries', European Economic Review, 88, pp. 158–184. doi:10.1016/j.euroecorev.2016.02.023.

Kollmann, R. et al. (2016) 'The post-crisis slump in the euro area and the US: Evidence from an estimated three-region DSGE model', European Economic Review, 88, pp. 21–41. doi:10.1016/j.euroecorev.2016.03.003.

Leeper, E.M., Plante, M. and Traum, N. (2010) 'Dynamics of fiscal financing in the United States', Journal of Econometrics, 156(2), pp. 304–321. doi:10.1016/j.jeconom.2009.11.001.

Leeper, E.M., Traum, N. and Walker, T.B. (2017) 'Clearing up the fiscal multiplier morass', American Economic Review, 107(8), pp. 2409–2454. doi:10.1257/aer.20111196.

Li, S.M. and Spencer, A.H. (2015) 'Effectiveness of the australian fiscal stimulus package: A DSGE analysis', Economic Record, 92(296), pp. 94–120. doi:10.1111/1475-4932.12224.

Mountford, A. and Uhlig, H. (2009) 'What are the effects of fiscal policy shocks?', Journal of Applied Econometrics, 24(6), pp. 960–992. doi:10.1002/jae.1079.

Mucka, Z. and Horvath, M. (2022) Fiscal policy matters: A new DSGE model for Slovakia, Rada pre rozpočtovú zodpovednosť. Available at: https://www.rrz.sk/en/fiscal-policy-matters-a-new-dsge-model-for-slovakia/ (Accessed: 24 May 2023).

Mumtaz, H. and Theodoridis, K. (2020) 'Fiscal policy shocks and stock prices in the United States', European Economic Review, 129, p. 103562. doi:10.1016/j.euroecorev.2020.103562.

Samimi, A.J. et al. (2017) 'The Impact of Fiscal Policy on Macroeconomic Variables: New Evidence from a DSGE Model', International Journal of Business and Developement Studies, 9(2), pp. 29–54. doi:10.22111/ijbds.2017.3703.

Sims, C.A. (2002) 'Solving Linear Rational Expectations Models', Computational Economics, 20(1/2), pp. 1–20. doi:10.1023/a:1020517101123.

Susan Yang, S.-C. (2005) 'Quantifying tax effects under policy foresight', Journal of Monetary Economics, 52(8), pp. 1557–1568. doi:10.1016/j.jmoneco.2004.09.003.

Takyi, P.O. and Leon-Gonzalez, R. (2020) ‘Macroeconomic impact of fiscal policy in Ghana: Analysis of an estimated DSGE model with financial exclusion’, Economic Analysis and Policy, 67, pp. 239–260. doi:10.1016/j.eap.2020.07.007.

Traum, N. and Yang, S.-C.S. (2015) ‘When does government debt crowd out investment?’, Journal of Applied Econometrics, 30(1), pp. 24–45. doi:10.1002/jae.2356.

Wang, C. et al. (2020) ‘Shadow banking activity and entrusted loans in a DSGE model of China’, The Manchester School, 89(5), pp. 445–469. doi:10.1111/manc.12319.

Wang, S.-L. (2021) ‘Fiscal stimulus in a high-debt economy? A DSGE analysis’, Economic Modelling, 98, pp. 118–135. doi:10.1016/j.econmod.2021.02.009.

Xiao, B., Fan, Y. and Guo, X. (2018) ‘Exploring the macroeconomic fluctuations under different environmental policies in China: A DSGE approach’, Energy Economics, 76, pp. 439–456. doi:10.1016/j.eneco.2018.10.028.

Zeman, J. (2017) ‘Fiscal Multipliers in the Slovak Economy: A DSGE Simulation’, Ekonomický časopis, 65(6), pp. 505–519.

Zubairy, S. (2014) ‘On fiscal multipliers: Estimates from a medium scale DSGE model’, International Economic Review, 55(1), pp. 169–195. doi:10.1111/iere.12045.

Chapter 2

Fiscal policy and the business cycle: An argument for non-linear policy rules

2.1 Introduction

The Great Recession and, especially, the Covid crisis have led to a revitalization of the interest in fiscal policy. Unlike its sibling, monetary policy, the fiscal policy tool repertoire offers various ways of interacting with households and the economy in a way that is only constrained by the government's budget constraint and the government's will to legislate. As such, policymakers have started increasingly stepping in during economic crises by releasing unprecedented stimuli packages, namely the American Recovery and Reinvestment (ARRA) in late 2009 and, recently, the Coronavirus Aid, Relief, and Economic Security Act (CARES Act) in 2020 in the US. At the same time, fiscal policy interactions are not limited to economic crises but are also frequently deployed during economic upturns, as in the Tax Cuts and Jobs Act in 2017. This begs the questions, "How effective is fiscal policy in stimulating the economy?", "Is stimulus more or less effective in deep recessions?" and "How do fiscal stimulus packages affect the economy in upturns?". This chapter aims to study these questions by exploring the dependency of fiscal policy effectiveness across the business cycle in a New Keynesian framework.

One of the main issues that arise when trying to estimate the effectiveness of fiscal policy is its endogeneity to the business cycle. The empirical Vector Autoregressive (VAR) literature has proposed various solutions from short-run restrictions and sign restrictions to the proxy structural VAR (SVAR) approach (see, for example, Blanchard and Perotti (2002), Mountford and Uhlig (2009), Mertens and Ravn (2014)). The findings in this literature differ greatly depending on the identification strategies, sample selection, and other factors.

More importantly, the standard approach relies on linear models due to a variety of reasons. Standard linear models estimate the average effect of fiscal policy across the business cycle.

However, it seems reasonable that the effect of a given policy intervention can vary depending on the state of the economy and its participants. To illustrate this, one scenario where one would expect fiscal policy to be more effective than normal is the situation that caused the revival of interest in fiscal policy in the first place: the Zero Lower Bound (ZLB). While the ZLB is the most prominent case of state-dependent fiscal policy effects, others arguably exist. For example, other variables may include credit market imperfections such as liquidity-constrained households, a high public debt level, the degree of economic slack, the state of the labour market, and more.

Studying any type of business cycle dependency requires economists to rethink the use of fully linear models. In essence, the conclusions that can be drawn from linear models are restricted to the average effect across the business cycle and do not necessarily give a full picture on the question: “Is stimulus more or less effective in deep recessions?”. Parker (2011, p. 708) puts this concisely:

> In the linearised model, the study of optimal fiscal policy is based on the answer to the question 'can the government raise model-based utility by conditioning government spending linearly on the state of the economy given that its effects are always the same?' and not 'can the government raise output or consumption more by increasing government spending in a recession than a boom and so should it?’

To explore this question, one must move away from linear models, which are unable to capture these higher-order, state-dependent dynamics. In fact, both the VAR literature (see, for example, Auerbach and Gorodnichenko (2012), Baum and Koster (2011), Ramey and Zubairy (2018), among others) and the DSGE literature have started exploring how fiscal policy actions may vary across time and economic scenarios. In the latter group, some of the notable studies that estimate state-dependent fiscal multipliers in different ways are Davig and Leeper (2010), Gomes et al. (2015), Sims and Wolff (2013), Sims and Wolff (2018a) and Sims and Wolff (2018b).

To shed light on how the effects of fiscal policy depend on the business cycle I follow the non-linear DSGE approach as in Sims and Wolff (2018a) and Amisano and Tristani (2010) and estimate the model on US data from 1984Q1 to 2021Q4. The core idea in Sims and Wolff (2018a) is that the structural equations of the DSGE entail useful information about how the effects of fiscal policy and fiscal policy itself relate to the measurements that characterize the business cycle. These types of effects cannot be captured by a first-order linearization, and therefore, I rely on a higher-order approximation. For this chapter, I develop a New Keynesian model with a rich fiscal and monetary ruleset that is closely related to Christiano et al. (2005) and Sims and Wolff (2018a) and shares significant similarities with Smets and Wouters (2007).

The fiscal ruleset includes several instruments such as consumption taxation, labour taxation, government consumption and transfers. The design of the fiscal instruments and their rules is based on Leeper, Plante and Traum (2010). Based on evidence for general non-linearity in the fiscal mechanism, as shown by Fernández-Villaverde et al. (2015), I include an alternative, non-linear component in the fiscal rules in the form of a restricted second-order Taylor approximation. The final rules can transmit state dependency but also generate business cycle dependency by themselves. This allows the government to vary its responses to the business cycle based on the current economic circumstances. Choosing a particularly general ruleset allows the data to speak expressively about the dynamics. For example, the government consumption variable may act differently depending on the state of the government's financial situation. If government debt is particularly high, it may be the case that government consumption expenditures have to be financed by relying more on tax hikes as opposed to raising debt. If the way the government acts changes based on the economic circumstances, then arguably, the effects of fiscal policy are likely to change, too. To illustrate this point, Leeper, Plante and Traum (2010) have shown that the adjustment speed to government debt is a fundamental determinant of the impact of fiscal policy. The model developed for this chapter can capture endogenous changes to the adjustment speed to debt and thus is able to predict a much wider range of possible effects for fiscal policy.

Furthermore, I include similar non-linearities in the monetary policy rule to partially capture the ZLB mechanics. Standard interest rate rules designed to capture the Zero Lower Bound mechanics feature a kink design, in which the interest rate is driven by a standard Taylor rule if the rate is above the ZLB and fixed at some low constant at the ZLB. Instead, I use a ruleset in which the Central Bank may vary its responsiveness to inflation and output growth in accordance with a second-order Taylor approximation. With the financial crisis in mind, it seems reasonable that the crisis caused a shift in the emphasis of the central banks from controlling inflation towards controlling output.

I estimate the model and show how the following two things depend on the initial conditions of the economy: the behaviour implied by fiscal policy rules and the impact of fiscal policy interventions on output.

Starting off with the effects of policy interventions, allowing impulse response functions to vary across business cycle conditions substantially increases the uncertainty about the effects of fiscal policy. This may explain why the empirical VAR literature generates such a broad range of results. I also find that all fiscal instruments are more expansionary in low-interest rate periods and, overall, less expansionary in periods of high debt, similar to Fotiou et al. (2020). The effects of government consumption are estimated to be countercyclical to output, while tax cuts are procyclical.

Combining the results on business cycle dependency of the effect of policy interventions with estimated business cycle conditions across US history from 1984Q1 to 2021Q4 allows me to trace out a timeline of the effectiveness of fiscal shocks. I find that government consumption goes through deep cycles, and it was substantially more effective during the financial crisis and the Covid crisis.

Moving on to how fiscal policy responds to the economy, I trace out how the responsiveness of fiscal variables output and debt changes across the sample. I show that most gradients respond to the debt level and adjust to ensure financial stability. For example, during the high debt period, which begins in the early 1990s, transfers and labour taxation start becoming more responsive to debt and act more strongly to reduce the deficit.

In a similar fashion to fiscal policy, the monetary policy rule is also allowed to vary across the business cycle. I show that the central bank changes its behaviour based on current output growth and shifts its focus in economic downturns from controlling inflation to controlling output and vice versa.

The final contribution of this chapter comes in the form of the empirical methods used. I estimate the higher-order DSGE model based on particle filter techniques to capture as much of the non-linear dynamics as possible. Estimating non-linear DSGE models is a computationally intensive exercise that is the main barrier preventing economists from using these models more regularly. Therefore, this chapter makes a particular effort to construct a sound methodology that trades off computation time and the quality of inference. Overall, the estimation time is reduced from weeks to days and depending on the comparison basis, computation time can be reduced by up to 94%. Moreover, I provide detailed guidelines for potential ways to cut down estimation time that I hope will be useful for others and will lead to wider use of the non-linear DSGE models.

This chapter is structured as follows. Section 2.2 presents a literature review. Section 2.3 sets up the model and presents the dynamic equations. Section 2.4 presents the estimation procedures employed to estimate the model, a detailed discussion on the construction of the data series with a particular focus on fiscal instruments, and an overview of the computational methodology I used for the estimation and posterior estimates. Section 2.5 presents the results on state dependency. The appendix includes more detail on the second-order pruned system, estimation diagnostics, posterior density plots, re-estimation results for Amisano and Tristani (2010) and more detail on the code implementation.

2.2 Literature review

2.2.1 VAR and linear DSGE models

Identifying and estimating the effects of fiscal policy intervention presents a series of complicated issues that have spawned a significant and diverse literature in macroeconomics. One of the main difficulties is the endogeneity problem of fiscal policy. Fiscal policy movements, as we can observe them, are typically not thought of as being purely exogenous.[10] Instead, fiscal policy action may, in part, be motivated by the business cycle at the time of intervention. This poses a problem because it becomes difficult to disentangle the effect of a policy intervention on output from the effects of automatic responses of fiscal policy to the business cycle.

The empirical VAR literature has produced a number of solution strategies to the identification problem, from imposing short-run restrictions, sign restrictions to the proxy SVAR approach. The canonical paper of Blanchard and Perotti (2002) imposes a mixture of short-run restrictions and outside calibration to identify exogenous movements. The key assumption is that fiscal policy

[10] Though it is argued that specific tax changes may be exogenous as for example in Romer and Romer (2010).

lags behind in its response to the business cycle. Their results show underwhelming effects of tax interventions, which are small on impact and fail to produce multipliers above one. Using a different approach, namely, sign restrictions, Mountford and Uhlig (2009) find that fiscal policy intervention can be highly effective with multipliers of up to three over a longer horizon. Similarly, Mertens and Ravn (2014) find higher fiscal multipliers in the short and medium run using a proxy SVAR approach that combines short-run restrictions with the narrative approach to identify effects.

The VAR literature explores fiscal multipliers in linear models, which also encompass the linear DSGE model category. Linear models assume that the effect of policy interventions is independent of the state of the economy and is identical in all economic circumstances. In other words, it is based on a study of the average effect. To account for the fact that policy intervention can have varying effects depending on the state of the economy (e.g. as a result of more binding credit constraints) and can itself be a function of the state of the economy (e.g. fiscal policy rules that depend on output or debt in a non-linear fashion), the literature has moved towards more flexible models.

On the VAR side, Auerbach and Gorodnichenko (2012) pioneered the use of regime-switching VAR models with smooth transitions. Regime-switching VAR models divide the business cycle into phases, and transitioning between phases may be induced by a set of economic circumstances. In each phase, the economy behaves according to a standard linear VAR model and is conditionally linear. The consequence is that fiscal policy effectiveness can vary from phase to phase. The results of Auerbach and Gorodnichenko (2012) established two key ideas. Firstly, they find that fiscal policy effectiveness does indeed vary across the phases. Secondly, they find strong evidence that fiscal policy behaves in the classical Keynesian sense. For expansionary phases, they find that the government spending multipliers are between 0 and 0.5 and in depressions or recessions, the multiplier is between 1 and 1.5.

Auerbach and Gorodnichenko (2012) spawned an entire literature on state-dependent effects of fiscal policy in VAR models. Baum and Koster (2011), Ferraresi, Roventini and Fagiolo (2014) and Fazzari, Morley and Panovska (2015) all find results consistent with the classical Keynesian worldview in that fiscal policy seems to be more effective in phases of negative output gaps, tight credit regimes and considerable economic slack, which are typically associated with economic downturns. However, there is also somewhat contradictive evidence provided by Ramey and Zubairy (2018) and Owyang, Ramey and Zubairy (2013). Both papers suggest that fiscal multipliers may not be as dependent on economic slack and do not generally deliver multipliers larger than unity. Ramey and Zubairy (2018) argue that the difference arises from different assumptions in the construction of the impulse responses. In particular, for the construction of the impulse responses, Auerbach and Gorodnichenko (2012) assume that the economy will stay in the initial state for 20 quarters, while Ramey and Zubairy (2018) aim to take into account the average duration of each phase. The uncertainty in state-dependent effects goes even further, as Arin et al. (2015) suggest that tax multipliers may even be procyclical.

More recently, in a Panel Vector Auto Regression model, Huidrom et al. (2020) find that there is a relationship between fiscal multipliers and fiscal positions. Their results show that fiscal multipliers are smaller when the fiscal positions are weak. Fotiou et al. (2020) find that the output effect of capital income tax cuts is dependent on government debt. Output multipliers become expansionary when debt is low and decrease in effectiveness when debt is high. Similarly, in a study focusing on debt stabilization, Fotiou (2022) finds that the initial conditions of government debt are determinants of the effects of fiscal policy interventions on output growth. They find that if government debt is low, then tax-based shocks are more productive on output growth in expansions than in recessions. Demirel (2021) shows that the effects of tax changes are more muted in periods of high unemployment.

Similar to the VAR literature, the fiscal DSGE literature has also emphasized business cycle dependency of the effects of interventions. The main focus so far has been on introducing specific mechanisms that allow fiscal policy effects to vary. In a seminal paper, Woodford (2011) explores how the effectiveness of government purchases varies with the type of monetary accommodation by the central bank in analytically tractable New Keynesian models. The central finding of Woodford (2011) is that when the central bank follows its targeting rule for the interest rate, then fiscal policy is less effective and can only offer multipliers of up to one. However, if monetary policy is constrained, fiscal multipliers become significantly more effective. Similar ideas were developed in Christiano, Eichenbaum and Rebelo (2011). Drautzburg and Uhlig (2015) and

Boubaker, Khuong Nguyen and Paltalidis (2018) provide empirical evidence by estimating DSGE models with Zero Lower Bound constraints, and they find consistent results.

A different mechanism that has been explored is the role of fiscal policy in heterogeneous agent models. While the representative household of an economy may not experience hard constraints in a crisis, sub-sets of the population may, for example, be credit-constrained. By definition, credit-constrained households are limited in their ability to borrow. What that means is that in a crisis, these households will not be able to borrow against future income to smooth consumption today in the same way as their Ricardian counterparts. Transfers, government consumption expenditures or tax cuts allow these households to avoid the hard credit constraints and to directly raise their consumption closer to the level of the Ricardian agents. Roeger and in't Veld (2009) show that an increased share of non-Ricardian households can increase the effectiveness of fiscal policy measures drastically. Furthermore, the introduction of Ricardian and non-Ricardian households introduces a natural source of variation for fiscal policy effectiveness across the business cycle, as explored in Krajewski and Szymansk (2019). They show that recessions can increase the share of non-Ricardian households, and as this share rises, fiscal policy becomes more effective. Other papers that focus on heterogenous agent models with credit constraint agents include Galí et al. (2007) and Kaplan and Violante (2014).

2.2.2 Non-linear DSGE models

Modern DSGE models are defined by a set of linear and non-linear equations. Typically, these models are not solvable in their general form, with the exception of simplistic models. In practice, one often resorts to varying levels of Taylor approximations or conditionally linear models.[11] Naturally, if one approximates a model, some of the original characteristics of the model may be lost. The key question here relates to how non-linear DSGE models are. As DSGE models feature

[11] Note that first-order Taylor approximations are the perfectly appropriate in many scenarios depending on the model, data and modelling framework. Further, they can have huge computational advantages when it comes to inference. On the question of how appropriate first-order Taylor approximations, the answer seems to be: it depends. In many smaller modelling frameworks there is evidence that linear models can have negligible approximations errors. However, there is also conflicting evidence that even in those cases. Fernández-Villaverde and Rubio-Ramirez (2005) and An and Schorfheide (2007) show that in small, typically, nearly linear models the effect of including higher-order terms can improve the fit of the models, change posterior distributions, and deliver different moment estimates.

linear or nearly linear equations, like the capital accumulation law, some subcomponents will necessarily behave approximately, if not exactly, linear. But frequently, economists introduce simple mechanics like multiplicative shocks and scale-dependent decision-making that can push a model to be more non-linear. To illustrate this, take a simple non-stochastic Euler Equation in a model with log utility:

$$\frac{1}{C_t} = \beta R_t \frac{1}{C_{t+1}}.$$

Euler equations define a trade-off between current, C_t, and future consumption, C_{t+1}, as governed by the real interest rate, R_t, and the discounting factor β. A standard question to ask would be, "What is the household's response to a change in the interest rate?" Here, I focus on the partial equilibrium case to build intuition on the problem of curvature. The answer to this question is it depends. The gradient of current consumption to interest rates reveals two things:

$$\frac{\partial C_t}{\partial R_t} = (-1)\beta C_t \frac{C_t}{C_{t+1}}.$$

Firstly, the partial equilibrium effect of an increase in interest rates implies a reduction in current consumption as all variables are positively valued. Secondly, the size of the reduction depends on the level of current and future consumption. For example, if future consumption is higher, then the gradient is smaller. If the Agent expects to be well-off in the future, there is less of an advantage to save, and thus, the Euler equation implies a smaller response to changes in the interest rate. Further, if the Agent is well-off today, it responds much stronger to changes in the interest rate and reduces consumption by more than if it was not well-off. Hence, even in this simple model, the consumption response to the changes in the interest rate is non-linear in the levels of both current and future consumption.

Sometimes, state dependency may be solved by a change of variable. For example, if one looks at the log of consumption and interest rate as the measure of interest, then the equation can be simplified as follows:

$$\frac{\partial \ln\,(C_t)}{\partial \ln\,(R_t)} = (-1).$$

However, other popular changes of variables like relative steady state deviations may not get rid of the state dependency without approximations:

$$\frac{\partial \tilde{C}_t}{\partial \widetilde{R}_t} = (-1)\beta R\Big(1 + \tilde{C}_t\Big)\frac{\Big(1 + \tilde{C}_t\Big)}{\Big(1 + \tilde{C}_{t+1}\Big)},$$

where non-index variables correspond to steady state values and $\tilde{x}_t$ corresponds to the percentage deviation from the steady state for the variable x_t. So, while in some incidences, state dependency can be solved, in general, it cannot be solved for all variable formulations and types of non-linear equations. For example, considering a more complex utility function with habit persistence would complicate things significantly. Consequently, by including higher-order approximation terms, we can learn about how the representative household may vary its response to economic variables depending on its circumstances.

Sims and Wolff (2018a) explore how fiscal policy effects of tax cuts may vary with business cycle conditions in a more general sense, where the properties of a higher-order Taylor approximation of the fiscal model are explored using parameter draws coming from a linear estimation of the same model. Sims and Wolff (2018a) focus on the co-movement between tax multipliers and the level of output in the business cycle. For a small and analytical example, they illustrate that labour tax cuts are state-dependent and, in particular, covary with the level of output and the level of taxation. In particular, they show that tax rate multipliers are larger when the level of output is higher, in contrast to classical Keynesian model predictions. For government spending, Sims and Wolff (2013) and Sims and Wolff (2018b) observe some variation but to a lesser degree.

In this chapter, I follow the approach by Sims and Wolff (2018a) by estimating a higher-order DSGE model with rich fiscal and monetary policy rules. Using the estimated model, I explore how the effects of fiscal policy interventions relate to the business cycle conditions and how the behaviour of the fiscal government, as implied by the fiscal rule functions, changes depending on the business cycle.

2.3 Model description

The following section describes the New Keynesian model developed in this chapter. The model is closely related to Amisano and Tristani (2010) but also features significant similarities with Smets and Wouters (2007) and Leeper, Plante and Traum (2010). The section is divided into four parts. Sections 2.3.1 and 2.3.2 describe the federal government and how the federal government rules change over the business cycle. Sections 2.3.3 and 2.3.4 do the same for the monetary rule set included in this model. Sections 2.3.5 , 2.3.6 and 2.3.7 complete the model setup by describing the household and firm problem followed by closing conditions. Lastly, section 2.3.8 describes the prior distribution of the model parameters.

2.3.1 Fiscal Policy

The key component of this model is its fiscal policy mechanism. Fiscal policy has become an increasingly important addition to the policy toolbox in crises. This is highlighted by large stimulus packages during the financial crisis of 2008 and during the Covid-19 crisis. As such, questions like “Is fiscal policy effective in economic crises?” or “How does the current state of the economy (in crisis or boom) affect the utility of fiscal policy?” are crucial and ought to be answered.

It is worth to note that even if the fiscal rules are linear, fiscal variables might respond to other variables such as consumption, output or federal debt, which in turn exhibit non-linear dynamics in the economy. However, I argue that allowing for non-linear fiscal policy rules significantly enriches the model for several reasons. First, it is reasonable to assume that governments follow different rulesets in financial crises than at and around the steady state. Second, it provides for more flexible response options for the fiscal variables than the standard model. To explore how the behaviour of fiscal policy changes, I later explore how the gradients of the fiscal policy rules change across the observed time period. The gradients in turn tell us something about the inner workings of the government and how it shifts the way it responds to the economy based on the state of the business cycle.

To illustrate the usefulness of rulesets that can vary across economic conditions, I will now delve into some scenarios where such rules may be advantageous. The standard way to model fiscal response functions is to constrain the debt and output response parameters in such a way that the government always responds to changes in debt and GDP to stabilize the budget. In practice, that implies government spending that is countercyclical to output and debt. In an economic downturn, the government is encouraged to start spending to bring the economy back on track and in upturns, it reduces spending to bring debt back to the steady state. For tax rates, the opposite applies. While mechanically a reasonable and desirable modelling property, there is evidence that government spending can, at times, be procyclical to output. Ideally, a ruleset would be able to represent both aspects of government spending. In addition, it can be argued that in severe economic crises, the government may choose to ignore or soften budgetary rules to stimulate the economy effectively. This can more easily explain how large financial packages like the American Recovery and Reinvestment Act or recent Covid measures are consistent with stable government dynamics.

To sum up, in order to capture the full potential range of business cycle dependency that fiscal policy offers, the ruleset is required to be flexible enough to vary across the cycle. In order to comply with this, I focus on the canonical fiscal rules design as in Leeper, Plante and Traum (2010). Their approach is to think about fiscal policy purely as a reaction function to its past values and the economy. Let z_t be a vector of fiscal variables and let $x_t$be a set of variables that fiscal policy responds to. This may include its past values, shocks, and other economic variables. The way fiscal policy responds is governed by a vector-valued function f that ought to be recovered. Together, fiscal policy can be defined as:

$$z_t = f(x_t).$$

In practice, the functional form of f is unknown. Leeper, Plante and Traum (2010), for example, assume that f is linear and fiscal instruments respond to past values of themselves, government debt and output. However, there are various ways to construct f depending on the *a priori* beliefs of the economist. To capture the two components of state dependency, I rely on a second-order Taylor approximation of f. This approximation is then restricted based on economic a-priori beliefs to build the final fiscal rules. Using Taylor approximations as fiscal rulesets has some advantages in the DSGE application. Firstly, the higher-order terms can allow the gradients of the response function to change across the business cycle and, thus, capture some of the desired dynamics. Secondly, as the Taylor approximation is smooth and unbounded, it easily integrates into the DSGE solution strategies.[12] A second-order Taylor approximation of f around the steady state, $\bar{x}$, can be constructed as follows:

$$z_t \approx \; f(\bar{x}) + Df(\bar{x})(x_t - \bar{x}) + \frac{1}{2}Hf(\bar{x}) * [(x_t - \bar{x}) \bigotimes (x_t - \bar{x})],$$

where $Df(\bar{x})$ is a matrix of first-order derivatives and $Hf(\bar{x})$ can be constructed based on the Hessian matrices of the individual equations. The remaining task is to restrict the different components of this approximation in a sensible way. Most common strategies rely on $Hf(\bar{x}) = 0$ in linear models and focus on $D\ f(\bar{x})$ and $f(\bar{x})$. The main feature utilized in this chapter is the option to parameterize $Hf(\bar{x})$ directly as it could deliver useful insights.

The model includes the following fiscal variables in the vector z_t at time t: consumption tax rate, τ_t^c, labour tax rate, τ_t^L, government consumption, G_t, and transfers, Z_t. This model excludes capital and, hence, capital taxation for the reason that computation time grows in a super-linear fashion with the number of states. The main restriction is that they respond linearly to their past values and shocks but may depend linearly and non-linearly on the economic variables of output, Y_t, inflation, π_t, productivity, A_t, and debt, B_t. The Taylor approximation can then be restricted to:

$$z_t = \; A + B(z_{t-1} - \bar{z}) + C(y_t - \bar{y}) + \frac{1}{2}D * [(y_t - \bar{y}) \bigotimes (y_t - \bar{y})] + Ev_t, \quad v_t{\sim}N(0, I),$$

where in this application:

$$z_t = [\tilde{\tau}_t^c, \tilde{\tau}_t^l, \tilde{Z}_t, \tilde{G}_t]' \text{ and } y_t = \left[\tilde{Y}_t, \tilde{\pi}_t, \tilde{A}_t, \widetilde{B}_t\right]'.$$

All variables are expressed in terms of steady state deviations, as indicated by the tilde. The matrix A of the approximation is set to 0, and $\bar{z}$ and $\bar{y}$ can be dropped because they are zero at the steady state. v_t is the vector of fiscal shocks. For this application, I set B and E to diagonal

[12] To illustrate this, an alternative could be a piecewise linear approach. Arguably, one could divide the fiscal system into two subsystems: one that is active in crisis and a standard reference system. However, this brings other challenges with it, like estimating which system is active when.

matrices with parameters along the diagonal. C and D are fully parameterized to capture the potential non-linearity of fiscal policy for all instruments but τ_t^c. Based on Leeper, Plante and Traum (2010), the federal consumption tax rate in the US focuses mainly on taxes for specific goods like gasoline or cigarettes. Because of this, the process for τ_t^c is restricted to be linear, exogenous and expressed in log steady state deviations:

$$\tilde{\tau}_t^c = p_{\tau^c}\tilde{\tau}_{t-1}^c + \sigma_{\tau^c} v_t^{\tau^c}, \qquad v_t^{\tau^c} \sim N(0{,}1).$$

Here, p_{τ^c} is an autoregressive parameter with $p_{\tau^c} \in (0{,}1)$ and σ_{τ^c} is the standard deviation of the structural consumption taxation shock $v_t^{\tau^c}$. For the remaining fiscal variables, the law of motion can be rewritten as follows for a fiscal instrument $\tilde{x}_t$:

$$\tilde{x}_t = p_x\tilde{x}_{t-1} + (1-p_x)\Big(k\mu_{x,Y}\tilde{Y}_t + \mu_{x,\pi}\tilde{\pi}_t + k\mu_{x,B}\widetilde{B}_t + \mu_{x,A}\tilde{A}_t + 0.5 * \varphi_{x,Y,Y}\tilde{Y}_t^2 + \varphi_{x,\pi,Y}\tilde{\pi}_t\tilde{Y}_t + \varphi_{x,A,Y}\tilde{Y}_t\tilde{A}_t + \varphi_{x,B,Y}\tilde{Y}_t\widetilde{B}_t + 0.5 * \varphi_{x,\pi,\pi}\tilde{\pi}^2{}_t + \varphi_{x,\pi,A}\tilde{\pi}_t\tilde{A}_t + \varphi_{x,\pi,B}\tilde{\pi}_t\widetilde{B}_t + 0.5 * \varphi_{x,B,B}\widetilde{B}_t^2 + \varphi_{x,B,A}\widetilde{B}_t\tilde{A}_t + \varphi_{x,A,A}A_t^2\Big) + \sigma_x v_t^x, \quad v_t^x \sim N(0{,}1),\ k = 1\ if\ \tilde{x}_t = \tilde{\tau}_t^l\ and\ else\ k = -1,$$

where $p_x \in (0{,}1)$, $\mu_{x,Y} > 0$ and $\mu_{x,B} > 0$. σ_x corresponds to the standard deviation of the structural fiscal shock v_t^x. The remaining parameters are unbounded. Thus, the fiscal instruments are allowed to respond to the changes in economic circumstances based on a particularly rich ruleset. The linear response terms govern the behaviour of the fiscal rule at the steady state of the economy, while state dependency is introduced via the inclusion of higher-order terms. As the economy moves away from the steady state, the second-order terms become active and may change the standard behaviour of the rules implied at the steady state.

To ensure the solvency of the federal government, it has to follow the budget constraint below. Based on the labour and consumption tax rates, it receives tax income on the corresponding tax bases of consumption expenditures, C_t, and total labour income, $\int_0^1 W_t(i)L_t(i)di$. Here $W_t(i)$ corresponds to the wage received by the household from firm i in a continuum of firms with a labour supply of $L_t(i)$. The government has expenditures in the form of transfers to households, Z_t, and government consumption expenditures, G_t. Lastly, the government gives out one-period bonds, B_t, to finance operations.

$$\frac{\tau_t^C}{1+\tau_t^C}C_t + \frac{\tau_t^L}{1+\tau_t^L}\frac{1}{P_t}\int_0^1 W_t(i)L_t(i)di + B_t = Z_t + G_t + \frac{I_{t-1}B_{t-1}}{\pi_t}.$$

2.3.2 State dependency of the fiscal rule set

To explore how fiscal policy may respond to the economy, I will now show how government consumption, $\tilde{G}_t$, responds to changes in federal debt in this ruleset as a representative case for the remaining variables. Differentiating the fiscal response function with respect to output, we get:

$$\frac{\partial \tilde{G}_t}{\partial \widetilde{B}_t} = (1 - p_G)\Big(-\mu_{G,B} + \varphi_{G,B,B}\widetilde{B}_t + \varphi_{G,A,B}\tilde{A}_t + \varphi_{G,Y,B}\tilde{Y}_t + \varphi_{G,\pi,B}\tilde{\pi}_t\Big).$$

If the economy is at its steady state, then marginal changes in the federal debt level, $\widetilde{B}_t$, have a fixed effect on $\tilde{G}_t$ as all other state variables are equal to zero and drop out. At the steady state, the responsiveness is governed by $(1 - p_G)(-\mu_{G,B})$ with $\mu_{G,B} > 0$ and $p_G \in (0,1)$. That means that as government debt goes up, government consumption goes down to stabilize the budget. So, for a given set of parameters, a linear ruleset and a second-order ruleset observed at the steady state are indistinguishable. However, as the economy moves away from the steady state, the gradient $\frac{\partial \tilde{G}_t}{\partial \widetilde{B}_t}$ may change linearly in inflation, output, productivity and government debt. Assuming $\varphi_{G,B,B} < 0$ as an example, then the gradient $\frac{\partial \tilde{G}_t}{\partial \widetilde{B}_t}$ is decreasing in the debt variable. As the federal debt level rises above the steady state, $\varphi_{G,B,B}\widetilde{B}_t$ becomes negative and reduces the overall gradient of government consumption to debt. This would, for example, be the case for a government that favours austerity policies. As the federal debt level increases, the government becomes more concerned with stabilizing the debt level and the responsiveness to debt increases in absolute terms. However, in low debt periods, $\varphi_{G,B,B}\widetilde{B}_t$ is positive and increases the gradient. In absolute terms, in low debt periods, government spending is then potentially less responsive to debt and may even become procyclical. The consequence is that the responsiveness of government consumption to federal debt is asymmetric.

2.3.3 Monetary policy

Next to the federal government, this model also features a central bank. The central bank operates based on a Taylor-like rule. Like the federal government, the central bank rule also features second-order terms:

$$\begin{aligned} i_t = (1-\rho_I)(\bar{\pi} - ln(\beta) + \psi_y(y_t - y_{t-1}) + \psi_\pi(\pi_t - \pi_t^*) + 0.5\psi_{y,y}(y_t - y_{t-1})^2 \\ + \psi_{y,\pi}(y_t - y_{t-1})(\pi_t - \pi_t^*) + 0.5\psi_{\pi,\pi}(\pi_t - \pi_t^*)^2) + \rho_I i_{t-1} + v_t^i, \qquad v_t^i \sim N(0, \sigma_i^2). \end{aligned}$$

The log interest rate today, i_t, responds autoregressively to last quarter's log interest rate i_{t-1} as governed by the AR(1) coefficient $\rho_I \in (0,1)$. Further, the rate responds to current output growth constructed as the difference between log output today and lagged log output, $(y_t - y_{t-1})$, and the difference between log inflation, π_t, and the log inflation target, π_t^*. To ensure stable inflation dynamics, ψ_π is larger than one and ψ_y is assumed to be larger than zero. v_t^i is the monetary policy shock. The higher-order parameters are unbounded. The log inflation target, π_t^*, follows a simple AR(1) process:

$$\pi_t^* = (1-\rho_\pi)\bar{\pi} + \rho_\pi \pi_{t-1}^* + v_t^\pi, \qquad v_t^\pi \sim N(0, \sigma_\pi^2).$$

Letting the inflation target vary across time gives the central bank some wiggle room with the way it responds to inflation. For example, if the economy faces inflationary pressure, then a Taylor rule dictates a rise in the interest rate. Here, the central bank may choose to relax the inflation target. As the inflation target increases, the overall response to inflation decreases, and the Taylor rule supports a slower return to the steady state. Vice versa, the central bank may choose to tighten its inflation target, forcing a quicker return to the steady state.

2.3.4 State dependency of the Monetary rule set

Similarly to the fiscal rule, the above Taylor rule behaves as a linear rule at the steady state:

$$\left.\frac{\partial i_t}{\partial(y_t - y_{t-1})}\right|_{steady\ state} = (1-\rho_I)\psi_y,$$

$$\left.\frac{\partial i_t}{\partial(\pi_t - \pi_t^*)}\right|_{steady\ state} = (1-\rho_I)\psi_\pi.$$

However, as the economy moves away from the steady state, the second-order terms begin to bite:

$$\frac{\partial i_t}{\partial(y_t - y_{t-1})} = (1-\rho_I)\left(\psi_y + \psi_{y,y}(y_t - y_{t-1}) + \psi_{y,\pi}(\pi_t - \pi_t^*)\right),$$
$$\frac{\partial i_t}{\partial(\pi_t - \pi_t^*)} = (1-\rho_I)\left(\psi_\pi + \psi_{y,\pi}(y_t - y_{t-1}) + \psi_{\pi,\pi}(\pi_t - \pi_t^*)\right).$$

Both gradients respond to both output growth and inflation above target. The parameter $\psi_{y,\pi}$ is shared by both gradients and, for example, governs the relationship of the gradient of the log interest rate to output growth with inflation above target. For example, assuming $\psi_{y,\pi} > 0$ implies that the focus on inflation, as implied by the gradient $\frac{\partial i_t}{\partial(\pi_t - \pi_t^*)}$, increases if the output growth rate is above zero. That implies that the central bank reacts stronger to the inflation rate being above target if the economy is in a boom phase. At the same time, the responsiveness to output growth may increase or decrease depending on whether inflation is above or below target, respectively. The two parameters $\psi_{y,y}$ and $\psi_{\pi,\pi}$ are not shared and, hence, describe one-sided effects. To illustrate, if $\psi_{y,y} > 0$, then the responsiveness of the interest rate to output growth is increasing in output growth or vice versa.

2.3.5 Household problem

This model features a very standard new Keynesian household problem, which builds on the Amisano and Tristani (2010) model. Here, the representative household optimizes the sum of discounted utility subject to a budget constraint as governed by the discount factor $\beta \in (0,1)$. The target function includes both consumption utility and labour disutility in an additively separable form. The agent derives utility from consumption, C_t, which is weighted against habit-adjusted lagged consumption, hC_{t-1}. The habit persistence is governed by the parameter $h \in (0,1)$, which is included to generate positive autocorrelation in consumption observed in the data. The deviation of consumption to last periods habit stock is weighted to the power of $1-\gamma$, where $\gamma > 1$ is a risk aversion parameter. Consequently, the utility function features diminishing marginal utility of consumption. The household also derives disutility from supplying labour to a continuum of firms. The household supplies labour, $L_t(i)$, in period t to firm i. The labour supply is differentiated to allow for Calvo pricing in the firm problem. In return for the labour supplied, the household receives a wage rate form firm i in the form of $W_t(i)$. As labour is supplied, the household receives disutility governed by parameters χ and ϕ. The agent integrates over the individual disutilities received from supplying work to all firms. The maximization problem is as follows:

$$\max_{C_t, L_t(i), B_t} E_0 \sum_{t=0}^{\infty} \beta^t \, U\big(C_t, C_{t-1}, L_t(i)\big), \qquad where$$

$$U\big(C_t, C_{t-1}, L_t(i)\big) = \frac{(C_t - hC_{t-1})^{1-\gamma}}{1-\gamma} - \int_0^1 \chi L_t(i)^{\phi} \, di$$

$$s.t. \left(1 + \frac{\tau_t^C}{1+\tau_t^C}\right) P_t C_t + P_t B_t = P_t Z_t + I_{t-1} P_{t-1} B_{t-1}$$

$$+ \left(1 - \frac{\tau_t^L}{1+\tau_t^L}\right) \int_0^1 W_t(i) L_t(i) di + \int_0^1 \Xi_t(i) di.$$

In the maximization problem, the household faces a budget constraint. The household receives funds in the form of labour income from the differentiated firms, $W_t(i)L_t(i)$, which are taxed by the federal government based on the labour taxation rate, τ_t^L. Further, the household receives government transfers, Z_t, and residual firm profits, $\Xi_t(i)$. At the same time, the agent has expenditures in the shape of nominal consumption expenditures, $P_t C_t$, where P_t is the current price level. The consumption expenditures are taxed based on the consumption tax rate, τ_t^C. Furthermore, the household has the ability to smooth consumption by purchasing government bonds, $P_t B_t$, today. At the same time, it pays interest on last periods bond holding, $I_{t-1} P_{t-1} B_{t-1}$, where I_{t-1} is last period's interest rate. The optimization problem leads to the following first-order conditions:

$$\frac{1}{1+\tau_t^l} \frac{W_t(i) L_t(i)}{P_t} = \frac{\chi \phi L_t(i)^{\phi}}{\Lambda_t},$$

$$\Lambda_t \left(1 + \frac{\tau_t^c}{1+\tau_t^c}\right) = (C_t - hC_{t-1})^{-\gamma} - \beta h E_t[(C_{t+1} - hC_t)^{-\gamma}],$$

$$\frac{1}{I_t} = \beta E_t \left[\frac{P_t}{P_{t+1}} \frac{\Lambda_{t+1}}{\Lambda_t}\right].$$

The first equation defines the trade-off between labour income and labour disutility, which is distorted by the labour taxation rate. The second equation is a standard consumption Euler equation. As such, it governs the trade-off between current and future consumption. However, here the key statistic is habit-adjusted consumption. The equation is distorted by the consumption tax rate. Λ_t corresponds to the nominally valued Lagrange multiplier of the constrained optimization problem. The last equation is the saving equation derived based on the preference for government bonds.

2.3.6 Firm problem

The following section lays out the firm sector of the model, which is comparable to Smets and Wouters (2007), Amisano and Tristani (2010) and Christiano et al. (2011). The firm sector includes two main components: a competitive final good firm and a continuum of intermediate good firms. The competitive final good firm bundles the differentiated output, $Y_t(i)$, of all individual firms $i \in (0,1)$ into a single product of the economy, Y_t. The intermediate outputs, $Y_t(i)$, are purchased from the continuum of intermediate firms. To do so, the final good firm uses a CES aggregator of the following design:

$$Y_t = \left(\int_0^1 Y_t(i)^{\frac{\theta-1}{\theta}} di \right)^{\frac{\theta}{\theta-1}}.$$

θ is the goods elasticity of substitution with $\theta > 1$. Each differentiated output, $Y_t(i)$, has a corresponding purchase price, $P_t(i)$, which the final firm takes as given. Based on this, the final good producers solve the following profit maximization problem:

$$\max_{Y_t, Y_t(i)} P_t Y_t - \int_0^1 P_t(i) Y_t(i) di \quad s.t. \;\; Y_t = \left(\int_0^1 Y_t(i)^{\frac{\theta-1}{\theta}} di \right)^{\frac{\theta}{\theta-1}}.$$

Solving the first-order conditions delivers the following demand schedule for each individual good i:

$$Y_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\theta} Y_t \quad \textit{for all } i.$$

The demand for each good produced by firm i is proportional to the overall market output, Y_t, but individually depends on a relative price rating comparing the individual product price, $P_t(i)$, the market price level, P_t, which is reweighted to the power of minus the good elasticity of substitution. As the individual price increases relative to the average market price, the demand for good i decreases.

The intermediate continuum of firms faces two types of problems: a basic production problem and a sequential pricing problem. For the production problem, the individual firms have the following production technology:

$$Y_t(i) = A_t L_t(i)^{\alpha},$$

$$log(A_t) = \rho_A \, log(A_{t-1}) + v_t^A, \qquad v_t^A \sim N(0, \sigma_A^2).$$

Each firm i transforms its labour supply, $L_t(i)$, into to the intermediate output, $Y_t(i)$. The production function features diminishing returns to labour with $\alpha < 1$. A_t is a common production technology that is governed by an AR(1) process in log terms with $\rho_A \in$ (0,1). v_t^A corresponds to the common structural technology shock. Based on this production technology, the individual intermediate firms choose their utilization of the labour supply by minimizing labour costs subject to meeting the market demand for the individual goods, ${Y_t}^D(i)$:

$$\min_{L_t(i)} W_t(i) L_t(i) \;\; s.t. \;\; Y_t(i) = A_t L_t(i)^{\alpha} \geq {Y_t}^D(i),$$

with the corresponding Lagrangian set up:

$$\max_{L_t(i)} \mathcal{L} = -W_t(i) L_t(i) + \lambda_t(i)(\, A_t L_t(i)^{\alpha} - {Y_t}^D(i)),$$

where $\lambda_t(i)$ is the Lagrangian multiplier. Solving the above optimization problem delivers the following identity:

$$\lambda_t(i) = \frac{W_t(i) L_t(i)}{\alpha A_t L_t(i)^{\alpha}}.$$

In this, the Lagrangian multiplier, $\lambda_t(i)$, also represents the marginal cost of increasing production by one unit, $MC_t(i)$, functioning as a shadow price. After pinning down the labour demand of firm i, the individual firms are faced with a pricing problem. The pricing problem features the canonical Calvo pricing mechanism in order to introduce price stickiness. The idea is that not all firms are fully able to adjust prices in a given period. Instead, firms may be in a situation where they cannot adjust prices based on a probability, $\zeta \in$ (0,1). As that is the case, optimal pricing requires firms to look forward and figure out what the consequence of choosing a price today is for today and the future. As such, firms choose a price by optimizing profits across the expected lifetime of that price. Like in Amisano and Tristani (2010), firms are not permanently stuck with a given price but receive a sub-optimal price update. The chosen price in period t, $P_t(i)$, is updated using steady state inflation, $\bar{\pi}$, and aggregate inflation, $\frac{P_{t+s-1}}{P_{t-1}}$, to obtain a period $t+s$ price. Arguably, this ensures that individual prices are updated in a reasonable way, even if firms are not able to update prices for significant periods of time. The design of the indexed prices is as follows:

$$P_{t+s}(i) = P_t(i)(\bar{\pi})^{1-l}\left(\frac{P_{t+s-1}}{P_{t-1}}\right)^l,$$

where $l \in (0{,}1)$ is an indexation parameter. The firms reoptimize the following Lagrangian to solve for an optimal reset price:

$$\underset{P_t(i)}{max}\, \mathcal{L} = E_t \sum_{s=0}^{\infty} \zeta^s \beta^s \frac{P_t}{P_{t+s}} \frac{\Lambda_{t+s}}{\Lambda_t} (P_{t+s}(i)Y_{t+s}(i) - TC_{t+s}(i)).$$

In s periods from the starting point of the optimization problem, firms have a probability of ζ^s to be still stuck with the update reset price. In every period, firms receive a profit stream, $P_{t+s}(i)Y_{t+s}(i) - TC_{t+s}(i)$, where the marginal cost function is generated based on the labour supply choice problem above. Firms discount these future profit streams using the common stochastic discount factor $Q_{t,t+s} = \beta^s \frac{P_t}{P_{t+s}} \frac{\Lambda_{t+s}}{\Lambda_t}$. It is assumed that all firms are identical except for their choice of price. As that is the case, all firms set the same optimal price. Solving the Lagrangian first-order system and substituting in the household labour supply condition delivers the following set of equations as in Amisano and Tristani (2010):

$$\Upsilon_{2,t} = \frac{\alpha(\theta-1)}{\phi\chi\theta}\left(\frac{1-\zeta\left(\bar{\Pi}^{1-l}\frac{\Pi_{t-1}^l}{\Pi_t}\right)^{1-\theta}}{1-\zeta}\right)^{1+\frac{\theta}{1-\theta}\frac{\phi}{a}} \Upsilon_{1,t},$$

$$\Upsilon_{2,t} = (1+\tau_t^l)\frac{A_t^{-\frac{\phi}{a}}}{\Lambda_t} Y_t^{\frac{\phi}{a}} + E_t\zeta\beta\frac{1}{\Pi_{t+1}}\frac{\Lambda_{t+1}}{\Lambda_t}\Upsilon_{2,t+1}\bar{\Pi}^{-(1-l)\theta\frac{\phi}{a}}\Pi_t^{-\theta\frac{\phi}{a}l}\Pi_{t+1}^{1+\theta\frac{\phi}{a}},$$

$$\Upsilon_{1,t} = Y_t + E_t\zeta\beta\frac{1}{\pi_{t+1}}\frac{\Lambda_{t+1}}{\Lambda_t}\Upsilon_{1,t+1}\bar{\Pi}^{(1-l)(1-\theta)}\Pi_t^{l(1-\theta)}\Pi_{t+1}^{\theta}.$$

Together these equations govern the dynamics of the Philips curve. Current inflation, $\mathit{\Pi}_t = \frac{\mathrm{P}_t}{\mathrm{P}_{t-1}}$, is implicitly defined as a function of past and future inflation, the markup, and the marginal cost function. The variables $\Upsilon_{1,t}$ and $\Upsilon_{2,t}$ are convenient summary variables in the representation of the Philips curve but do not carry their own easily interpretable meaning. It is important to note that the equation system above is a generalization of the standard linear New Keynesian Philips curve. If one was to construct a first-order approximation, the canonical curve could be recovered. However, the system above features a more elaborate dynamic for inflation. As is the case for the fiscal equations, the scale of the inflation response implied by this Philips curve system depends on the exact business cycle conditions and may vary across the cycle.

2.3.7 Model solution and set up

The equations above govern the main dynamics of the DSGE together with a simple market clearing condition:

$$Y_t = C_t + G_t.$$

However, the market closing condition is used to substitute out consumption to avoid keeping track of additional variables. The model features seven purely exogenous processes for seven data series: $\widetilde{A}_t$, $\tilde{\tau}_t^c$, $\tilde{\pi}_t^*$, v_t^i, v_t^{tl}, v_t^Z and v_t^G where variables with a tilde are measured in log steady state deviations. The shock processes for the interest rate and fiscal variables do not depend on other variables. To complete the model, identity equations are added for variables that can be both pre-determined and endogenous depending on the lag (for example, government debt in the last quarter is pre-determined today, while government debt today is endogenous today). The resulting state vector, x_t, and the endogenous vector, y_t, then govern the system:

$$x_t = \left[\tilde{\pi}_{t-1}, \tilde{Y}_{t-1}, \tilde{\imath}_{t-1}, \widetilde{B}_{t-1}, \tilde{\tau}_t^l, \tilde{Z}_t, \tilde{G}_t, \tilde{A}_t, \tilde{\tau}_t^c, \tilde{\pi}_t^*, v_t^i, v_t^{tl}, v_t^Z, v_t^G\right]',$$
$$y_t = \left[\ \widetilde{\Upsilon}_{1,t}, \widetilde{\Upsilon}_{2,t}, \tilde{\pi}_t, \tilde{\imath}_t, \tilde{Y}_{t,} \tilde{A}_t, \widetilde{B}_t, \tilde{\tau}_t^l, \tilde{Z}_t, \tilde{G}_t\right]'.$$

2.3.8 Prior

The following section describes the prior distribution setup for the above-described model. As the model is closely related to the Amisano and Tristani (2010) model, a lot of parameters, in particular the core economic parameters, receive similar or related priors. However, some prior were adjusted for empirical performance to adjust to US data and to create similarities with other implementations. For a full summary of the priors, see Table 2.1 and Table 2.2.

The discount factor, β, receives a *Beta* prior with a mean of 0.995. This corresponds to an annual real rate of two per cent. In comparison to other papers, β is slightly higher, reflecting a significant share of post-2000s observations. For the following parameters, including the coefficient of relative risk aversion, habit persistence, disutility of labour and goods elasticity of substitution, the priors are as in Amisano and Tristani (2010). The price indexation and Calvo pricing parameters have received adjusted *Beta* priors with a mean of 0.5 and a standard deviation of 0.1 to be more in line with Sims and Wolff (2018a) and Smets and Wouters (2007). Further, the linear output growth coefficient in the interest rule receives a *Gamma* prior with a mean of 0.125 and standard deviation of 0.035. In comparison to Amisano and Tristani (2010), the mean is slightly higher and closer to Sims and Wolff (2018a) and Smets and Wouters (2007).

Table 2.1: Prior distributions for core model parameters

para	prior	mean	sd.	para	prior	mean	sd.
β	$Beta$	0.99500	0.00100	p_π	$Beta$	0.90000	0.09000
$\gamma-1$	$Gamma$	1.00000	0.70000	σ_{τ^l}	$Gamma$	0.04000	0.01000
h	$Beta$	0.70000	0.13800	σ_{τ^c}	$Gamma$	0.04000	0.01000
$\phi-1$	$Gamma$	3.00000	1.00000	σ_Z	$Gamma$	0.04000	0.01000
$\theta-1$	$Gamma$	7.00000	2.64500	σ_G	$Gamma$	0.04000	0.01000
ζ	$Beta$	0.50000	0.10000	σ_a	$Gamma$	0.04000	0.01000
l	$Beta$	0.50000	0.10000	σ_i	$Gamma$	0.00400	0.00100
$\psi_\pi-1$	$Gamma$	1.00000	0.18200	σ_π	$Gamma$	0.00125	0.00056
ψ_y	$Gamma$	0.12500	0.03500	τ^l	$Beta$	0.23000	0.00100
p_{τ^l}	$Beta$	0.90000	0.09000	τ^c	$Beta$	0.01500	0.00100
p_{τ^c}	$Beta$	0.90000	0.09000	s_g	$Beta$	0.06000	0.00100
p_Z	$Beta$	0.90000	0.09000	s_b	$Beta$	0.50000	0.01000
p_G	$Beta$	0.90000	0.09000	π	$Gamma$	0.00560	0.00020
p_a	$Beta$	0.90000	0.09000				
p_i	$Normal$	0.80000	0.10000				

Notes: The table presents the prior distributions for the core model parameters, autoregressive, shock and steady state parameters.

Moving on from the core economic parameters, almost all autoregressive coefficients receive a standard *Beta* prior with a mean of 0.9 and standard deviation of 0.09. The choice of *Beta* prior ensures that the autoregressive coefficients remain lower than one, and consequently, this ensures sufficiently stable eigenvalues. The only exception is the interest rate rule parameter, p_i, which receives a normal prior with a mean of 0.8. The shock standard deviation parameters almost all receive a *Gamma* prior with a mean of 0.04. Two exceptions are the standard deviation for the interest rate and the inflation target shock, which are adjusted downwards for empirical performance.

The model features several parameters that define steady state relationships. These priors are constructed using frequentist, long-run sample estimates. For example, priors for τ^l and τ^c are calibrated using the average tax rates over the sample. Further, the debt and government consumption to output ratios receive the same treatment. Lastly, the steady inflation rate receives a *Gamma* prior with a mean of 0.0056.

The linear fiscal parameters govern the mechanics of the federal government at the steady state. At the steady state, it is assumed that fiscal rules focus on debt sustainability. That is, if the debt rises, then expenditures are reduced, and taxes are raised. At the same time, if output rises, taxes are increased, and expenditures are reduced. Therefore, the respective parameters receive *Gamma* priors which, in combination with the signs in the fiscal rules, create the above-described behaviour. For the inflation and productivity response parameters, the choice of prior fell on an unassuming *Normal* prior with a mean of zero and a standard deviation of 0.1. Once the economy starts moving away from the steady state, the non-linear terms become active and start influencing the fiscal policy rules. All non-linear fiscal policy parameters receive a *Normal* priors with a mean of zero and a standard deviation of 0.2. The prior reflects agnostic believes about the interaction terms but is comparatively diffuse and can allow the data to speak for itself. The last group of parameters are the non-linear interest rate rule parameters, which receive a similar *Normal* prior with a mean of zero and a standard deviation of one.

Two parameters are fixed as in Amisano and Tristani (2010):

$$\alpha = 0.76 \ and \ \chi = 0.273.$$

To ensure convergence of the particle filter, the measurement equation includes measurement errors. Measurement errors need to be included in particle filter applications as they are used to smooth the likelihood and can help prevent particle impoverishment. In this application, I dogmatically set the measurement error to 20% of the standard deviation of the data series, as in Herbst and Schorfheide (2016).[13] The 20% threshold is somewhat ad-hoc and based on an empirical necessity for the estimation to run smoothly. Directly estimating the measurement errors using a full covariance matrix would be a more sophisticated approach. However, the model estimated in this chapter already features a large number of parameters and requires a significant amount of computational resources.

[13] Both linear and non-linear filters typically compare actual observations to predicted observations in some fashion. In the linear Kalman filter, the main requirement for this comparison to work is that the 1-step-ahead error is non-degenerate which is typically the case if there are more structural shocks than data series. However, roughly speaking, as particle filters compare observations and predicted observations conditioned on the particles themselves, any state uncertainty that may be present in the Kalman filter disappears in the particle filter. Without further adjustments, the comparison between observations and predicted observations becomes degenerate: There may only be one particle that can predict the observations exactly, but it is unlikely for this particle to ever be sampled. Therefore, it is common practice to include measurement error as a band-aid in models estimated using particle filters. The measurement error ensures that there is residual uncertainty between observations and predictions and that the "relative fit" density is well defined.

Table 2.2: Response function priors

para	prior	mean	sd.
$\mu_{\tau^l,Y}$	*Gamma*	0.15000	0.10000
$\mu_{\tau^l,B}$	*Gamma*	0.15000	0.10000
$\mu_{\tau^l,\pi}$	*Normal*	0.00000	0.10000
$\mu_{\tau^l,A}$	*Normal*	0.00000	0.10000
$\mu_{Z,Y}$	*Gamma*	0.15000	0.10000
$\mu_{Z,B}$	*Gamma*	0.15000	0.10000
$\mu_{Z,\pi}$	*Normal*	0.00000	0.10000
$\mu_{Z,A}$	*Normal*	0.00000	0.10000
$\mu_{G,Y}$	*Gamma*	0.15000	0.10000
$\mu_{G,B}$	*Gamma*	0.15000	0.10000
$\mu_{G,\pi}$	*Normal*	0.00000	0.10000
$\mu_{G,A}$	*Normal*	0.00000	0.10000
$\varphi_{i,j,k}$	*Normal*	0.00000	0.20000
$\psi_{i,j}$	*Normal*	0.00000	1.00000

Notes: This presents the prior distribution set up for the linear and non-linear fiscal rule parameters. In addition, it also includes the non-linear interest rate parameters.

2.4 Estimation procedure

2.4.1 Likelihood construction

As the model is to be taken to the data, the likelihood of the model needs to be constructed for the data set. This requires constructing the different posterior model state distributions across time, and based on these distributions, the likelihood can be evaluated. In linear Gaussian state space models, the construction is comparatively straightforward as one can construct the sequence of distributions analytically using the Kalman filter recursions. The existence of analytical expressions for the individual distributions provides several advantages: fast likelihood evaluations, relatively robust simulations and the Kalman filter recursions are straightforward to implement. In the application of this chapter, the model to be estimated is non-linear as it features the second-order terms of the Taylor approximation of the DSGE. Therefore, likelihood evaluations using the Kalman filter are not appropriate as they would not inform the sampler about the contributions of the higher-order terms on the fit of the model. An alternative is provided by using simulation-based filters, also called particle filters.

Particle filters approximate the posterior state densities using a set of sampled measurement points called particles. Roughly speaking, in any time period, one starts with an initial distribution of particles. Within the time period, the set of particles is propagated forwards using some transition density. The propagated particles and the implied observational vector are compared to actual data. This is followed by a resampling step, in which ill-fitting particles are discarded, and better-fitting particles are used to repopulate the set of particles.

A variety of particle filters have been proposed with varying success depending on the exercise at hand. The crucial choice in particle filtering is the mechanism by which the particles are propagated forward. If the proposal mechanism is well-tailored, then particles are sampled, which explain the data well. In that case, fewer particles have to be discarded. If, however, the proposal mechanism struggles to produce well-fitting particles, this can lead to ill-fitting approximations of the likelihood via particle impoverishment. In the canonical particle filter, the bootstrap particle filter, the position of new particles is proposed via forwards iterating the model equations. As observed by Herbst and Schorfheide (2016), this can be quite inefficient depending on the application and can require increasingly large sample sizes to ensure accurate likelihood evaluations.

One way to combat this comes by adapting the proposal distribution to current observations. If one can find a well-adapted density, then particles can be sampled that explain the data well, and consequently, a smaller share of particles has to be discarded. In this application, I rely on the particle filter proposed by Amisano and Tristani (2010): the conditional particle filter. The filter linearizes the measurement equations of the DSGE. Based on the linearized measurement equation, one can sample particles using a conditional Gaussian density, just like in the Kalman filter. Based on testing for this chapter, the particles are sampled from well-adapted densities that only struggle with highly unlikely events like the Covid crisis. As a downside, the conditional particle filter abstracts away the non-linearity generated in the measurement equation. If the measurement equation happens to be very non-linear, then the approximation may be quite inaccurate. In terms of general performance, an analysis by Yang and Wang (2015) shows that the conditional particle filter outperforms the canonical filter by a wide margin and consequently requires significantly fewer particles (40 or more times fewer particles).

In the following, I first describe the Gomme and Klein (2011) DSGE solution system around which the conditional particle filter is built around. Based on this, I explore the main components of the Amisano and Tristani (2010) filter. The second-order approximation in the Gomme and Klein (2011) sense contains two transition systems: one system for the predetermined state variables and one system for the non-predetermined, endogenous variables. For the state variable vector, x_{t+1}, the system is governed by the following law of motion:

$$x_{t+1} = 0.5 * h_{\sigma\sigma} + H_x x_t + 0.5 * H_{xx}(x_t \bigotimes x_t) + \sigma J v_{t+1}, \qquad v_{t+1} \sim N(0, I),$$

where x_{t+1} is a vector of size $(n_x \times 1)$. $h_{\sigma\sigma}$, H_x and H_{xx} are matrices of sizes $(n_x \times 1)$, $(n_x \times n_x)$ and $(n_x \times n_x^2)$, respectively. v_{t+1} is a vector of size $(n_v \times 1)$ and corresponds to the structural, identified shock vector. J is a $(n_x \times n_v)$ matrix that governs the within-period impact of the structural shocks on the state vector. σ is the perturbation scalar and is typically set to one. The main dynamics of the DSGE are generated and propagated by the above system. The behaviour of the non-predetermined and endogenous variables stacked into a $(n_y \times 1)$ vector, y_{t+1}, is governed by the following system:

$$y_{t+1} = 0.5 * g_{\sigma\sigma} + G_x x_{t+1} + 0.5 * G_{xx}(x_{t+1} \bigotimes x_{t+1}),$$

where $g_{\sigma\sigma}$, G_x and G_{xx} are matrices of sizes $(n_y \times 1)$, $(n_y \times n_y)$ and $(n_y \times n_y^2)$. Notably, the system for y_{t+1} does not include autoregressive components and purely depends on the distribution of x_{t+1}. Furthermore, one may connect the variables contained in y_{t+1} to observables using a measurement equation of the following design:

$$y_{t+1}^{obs} = A + B y_{t+1} + e_{t+1}, \qquad e_{t+1} \sim N(0, \Sigma).$$

A and B are $(n_{obs} \times 1)$ and $(n_{obs} \times n_{obs})$ sized matrices, where n_{obs} is the number of observables and the row dimension of the vector y_{t+1}^{obs}. Further, the model includes a measurement error, e_{t+1}, assumed to be of size $(n_{obs} \times 1)$. The measurement error is normally distributed with covariance matrix, Σ. A convenient thing to do is to only keep track of the vector y_{t+1}^{obs}, and not of y_{t+1}^{obs} and y_{t+1}. The advantage of doing so comes in the form of a system reduction because typically $n_y > n_{obs}$. This can cut down the simulation time. To do so, one can substitute out y_{t+1} in the following way:

$$y_{t+1}^{obs} = A + B\Big(0.5 * g_{\sigma\sigma} + G_x x_{t+1} + 0.5 * G_{xx}(x_{t+1} \bigotimes x_{t+1})\Big) + e_{t+1},$$

$$y_{t+1}^{obs} = 0.5 * g_{\sigma\sigma}^{obs} + G_x^{obs} x_{t+1} + 0.5 * G_{xx}^{obs}(x_{t+1} \bigotimes x_{t+1})) + e_{t+1},$$

where the new matrices $g_{\sigma\sigma}^{obs}$, G_x^{obs} and G_{xx}^{obs} are of sizes $(n_{obs} \times 1)$, $(n_{obs} \times n_x)$ and $(n_{obs} \times n_x^2)$. The equation of the state vector, in addition to this new observational equation above, governs the system dynamics. This concludes the system description, and now I will provide a quick and condensed overview of the conditional particle filter as in Amisano and Tristani (2010).

Suppose one has a particle system of N draws from the time t distribution of the model states x_t for a given structural parameter vector of the DSGE, θ. Each particle is indexed using i as $x_{i,t}$. Then, the conditional particle filter relies on the following recursion to construct the likelihood:

1. Propagation step
    1.1. $x_{i,t+1} \sim p(x_{t+1}|y_{t+1}^{obs}, x_{i,t}, \theta)$ *for all* $i = 1,2,..,N$
2. Weight update step
    2.1. $w_i(x_{i,t+1}) = p(y_{t+1}^{obs}|x_{i,t}, \theta)$
3. Resampling step
    3.1. Resample the draws $x_{i,t+1}$ based on the importance weights $W_i = \frac{w_i(x_{i,t+1})}{\sum_{i=1}^{N} w_i(x_{i,t+1})}$

In the first step, the particles, $x_{i,t}$, are propagated forwards using a density p that is adapted to the current observational vector, y_{t+1}^{obs}. In the second step, weights are constructed using a measurement equation and based on those weights, the particles are resampled in the last step.

At its core, for the propagation step, the Amisano and Tristani (2010) filter relies on a linearization of the measurement equation around the expected value of x_{t+1}. Based on a linear measurement equation, one can construct a Gaussian density for x_{t+1} conditioned on the current observational vector. The first step lies in the construction of the expected value of x_{t+1}:

$$\bar{x}_{t+1|t} \approx \sum_{i=1}^{N} E(x_{i,t+1}|x_{i,t}) = \sum_{i=1}^{N} x_{i,t+1|t} = \sum_{i=1}^{N} 0.5 * h_{\sigma\sigma} + H_{\sigma} x_{i,t} + 0.5 * H_{\sigma\sigma}(x_{i,t} \bigotimes x_{i,t}).$$

The expectation $\bar{x}_{t+1|t}$ can be approximated via forwards iterating the individual particles $x_{i,t}$ for all N draws and setting the structural shocks to zero. The result is individual one-step-ahead predictions for the individual particles,$E(x_{i,t+1}|x_{i,t})$. $\bar{x}_{t+1|t}$ is obtained via averaging across the particle swarm. The second step is to generate a linearization of the measurement equation around $\bar{x}_{t+1|t}$ using the vector $x_{t+1|t}$. The linearization is of the format:

$$y_{t+1}^{obs} \approx y_{t+1|t}^{obs} + w_{t+1|t},$$

where the actual observable vector, y_{t+1}^{obs}, is approximately equal to some mean component, $y_{t+1|t}^{obs}$, and a new adapted measurement error, $w_{t+1|t}$. The component $y_{t+1|t}^{obs}$ is constructed as follows:

$$y_{t+1|t}^{obs} = 0.5 * g_{\sigma\sigma}^{obs} + (G_x^{obs} + 0.5 * G_{xx}^{obs} \overline{D}_k)\ x_{t+1|t} + 0.5 * G_{xx}^{obs}((x_{t+1|t} \bigotimes x_{t+1|t}) - \overline{D}_k \bar{x}_{t+1|t}),$$
$$where\ \ \overline{D}_k = \left[\frac{\partial(x_{t+1} \bigotimes x_{t+1})}{\partial x_{t+1}}\right]_{x_{t+1} = \bar{x}_{t+1|t}}.$$

Furthermore, $w_{t+1|t}$ is constructed as:

$$w_{t+1|t} = \sigma \bar{G}_x J v_{t+1} + w_{t+1},$$
$$where\ w_{t+1|t} \sim N\big(0, \sigma^2 \bar{G}_x J J' \bar{G}_x{}' +\ \Sigma\big)\ and\ \bar{G}_x = G_x + 0.5 * G_{xx}^{n} \overline{D}_k.$$

The approximated measurement equation has two important features: linearity and normality. Based on these two properties, one can construct a density for x_{t+1} conditioned on x_t and y_{t+1}^{obs}:

$$p(x_{t+1}|x_t, y_{t+1}^{obs}, \theta) \approx N(E(x_{t+1}|x_t, y_{t+1}^{obs}, \theta), V(x_{t+1}|x_t, y_{t+1}^{obs}, \theta))$$

This density can be used to effectively sample particles for x_{t+1} for the propagation step of the filter. In an approximate step, the weights, $w_i(x_{i,t+1})$, are constructed using the linearized measurement equation. This concludes the summary of the main components of the filter. However, there are some nuances for which I refer the reader to Amisano and Tristani (2010) for a more detailed and complete discussion. For the simulation, I utilize an initialization strategy based on Guerrieri and Iacoviello (2015). They use the first 20 observations as burn-in using a simpler filter to ensure that their non-linear filter starts from a well-adapted initial distribution. I follow this approach and use the Kalman filter for the first 20 observations. The number of particles for the conditional particle filter is set to 10,000.

2.4.2 Posterior simulation

Particle filters belong to a more general category of Sequential Monte Carlo (SMC) sampler. To be precise, particle filters are SMC samplers that are designed for state filtering and estimation. However, SMC samplers may also be applied to the estimation of Bayesian posterior distributions in the Del Moral, Doucet and Jasra (2006) sense. The following section first details why SMC samplers were chosen for this estimation and then describes the specific sampler used in this chapter.

SMC samplers have several advantages over other Bayesian simulation strategies typically used in the DSGE literature, specifically Markov Chain Monte Carlo techniques. Firstly, unlike Markov Chain samplers like the Random Walk Metropolis Hasting algorithm, basic SMC samplers can make effective use of multi-core CPUs. In the basic Random Walk Metropolis Hasting algorithm, every single likelihood has to be evaluated in sequence, while for SMC samplers, all likelihoods in the current particle system can be evaluated at the same time. For a fixed number of likelihood evaluations, this can provide immense computational gains inversely proportional to the number of cores in a CPU. Secondly, SMC samplers can be designed in a very adaptive manner and can, therefore, face a difficult trade-off between estimation accuracy and estimation time in an effective way. For a brilliant implementation of adaptive SMC samplers, see Buchholz, Chopin and Jacob (2021), which has significantly informed the SMC design in this chapter.

Moving on to the design of the SMC algorithm, SMC procedures divide the posterior estimation problem into a sequence of individually simpler estimation problems. To do so, one constructs a series that starts at an initial target density $\pi_1(\theta)$ for the structural parameter vector θ. The main quality of $\pi_1(\theta)$ is that it can be well approximated using importance sampling based on some initial proposal distribution, $\eta_1(\theta)$, used to populate the particle system. Once $\pi_1(\theta)$ has been approximated, the current sample can be used to approximate the next density in the series, $\pi_2(\theta)$. Ideally, if $\pi_1(\theta)$ and $\pi_2(\theta)$ can be chosen in such a way that they represent fairly similar densities, then the second approximation may also succeed. After the approximation, one can apply Markov Chain Monte Carlo steps to the individual particles to adapt them to $\pi_2(\theta)$ and to reintroduce variation lost in the sampling step. Iterating the two steps of importance sampling and adaptation allows for the construction of a series of distributions, $\{\pi_i(\theta)\}_{i=1}^{p}$, from $\pi_1(\theta)$ to $\pi_p(\theta)$ where $\pi_p(\theta)$ is the distribution of interest. In DSGE modelling, $\pi_p(\theta)$ that would be the posterior parameter distribution. SMC sampling has been utilized widely for parameter estimation in various models but has also been previously used in the DSGE literature in Herbst and Schorfheide (2016) and Creal (2007).

To construct an SMC algorithm, the crucial choice is the sequence of distributions. Based on the proposal in Del Moral, Doucet and Jasra (2006), I choose the following type of path:

$$\pi_n(\theta) \propto \pi(\theta)^{\phi_n} \mu_1(\theta)^{1-\phi_n},$$

with $0 \leq \phi_1 < \cdots < \phi_p = 1$. In the initial period one, the N particles $\{\theta_1^j\}$ are initialized based on some analytically tractable density $\mu_1(\theta)$ so that $\eta_1(\theta) = \mu_1(\theta)$. Further, for $\phi_1 = 0$ the initial target density is just $\mu_1(\theta)$. As ϕ increases, the sampler moves from the convenient $\mu_1(\theta)$ to the posterior, $\pi(\theta)$, as the weight of the initial proposal distribution decreases. This type of approach includes the Herbst and Schorfheide (2016) approach, which set $\mu_1(\theta)$ to the prior distribution of θ, $p(\theta)$. In that case:

$$\pi_n(\theta) \propto \big(p(y|\theta)p(\theta)\big)^{\phi_n} p(\theta)^{1-\phi_n} = p(y|\theta)^{\phi_n} p(\theta),$$

where information about the likelihood, $p(y|\theta)$, is added slowly to the prior. In this application, I follow the approach in Creal (2007) to use an initial distribution that approximates the target density, $\pi(\theta)$:

$$\pi_n(\theta) \propto \big(p(y|\theta)p(\theta)\big)^{\phi_n} \mu_1(\theta)^{1-\phi_n}.$$

This type of strategy is frequently applied to Random Walk Metropolis-Hastings samplers, which use an approximated posterior as the proposal distribution. Furthermore, it has some convincing advantages. While priors for structural economic parameters are informative about coverage of the parameters, in a lot of applications, one may find that prior and posterior beliefs for parameters differ substantially. In that case, the sampler spends significant time transversing vast low-density areas until it reaches an area of high likelihood. This is inconvenient both from a computational perspective and from an inference perspective. If one has an approximation of the posterior available, one can ensure that most of the likelihood evaluations take place in high-density areas. Ideally, if the approximation was perfect, $\mu_1(\theta) = p(y|\theta)p(\theta)$, then one could directly sample from the posterior. In reality, the sampler will correct a mismatch between the approximated posterior and the actual posterior. In this application, $\mu_1(\theta)$ is constructed by conducting 20 mode searches on an approximated posterior based on an estimation using the unscented Kalman filter. Likelihood evaluations using the unscented Kalman filter cost a fraction of the time and, therefore, are a convenient choice for the construction of $\mu_1(\theta)$. $\mu_1(\theta)$ is then constructed as a Gaussian centred at the mode with a diagonal approximation of the inverse Hessian scaled by factor two. The rescaling is done to ensure sufficient coverage.

The incremental weights of the SMC sampler can be defined as follows:

$$\tilde{w}_n^j(\theta_n^j, \theta_{n-1}^j) = \ (p(y|\theta_{n-1}^j)p(\theta_{n-1}^j))^{\phi_n - \phi_{n-1}} \mu_1(\theta_{n-1}^j)^{\phi_{n-1} - \phi_n} \propto \frac{\pi_n(\theta_{n-1}^j)}{\pi_{n-1}(\theta_{n-1}^j)},$$

where θ_{n-1}^j is the draw j of θ in iteration $n-1$ and $\tilde{w}_n^j$ is the incremental weight in period n. The incremental weights can be defined as above due to the choice of target densities and as I utilize invariant MCMC steps for the mutation step. Furthermore, based on the incremental weights, the normalized importance weights can be constructed as follows:

$$W_n^j = \frac{W_{n-1}^j \tilde{w}_n^j}{\sum_{j=1}^N W_{n-1}^j \tilde{w}_n^j} \ ,$$

where W_n^j is the normalized importance weights for particle j in iteration n. To finalise the sequence of distribution, one has to choose the sequence of temperatures governed by ϕ_n. A good tempering schedule ensures that all bridge distributions are always close enough to provide effective approximations. However, equally important is that the bridge distributions are not too close to each other. If they are too similar, the algorithm will spend significant time approximating perhaps virtually indistinguishable densities. Designing a good schedule is not a trivial problem and could require expensive test runs to get the tempering schedule right. For this application, I developed code for an adaptive tempering procedure created by Jasra et al. (2010). Jasra et al. (2010) rely on the effective sample size (*ESS*) to adaptively construct the tempering schedule for the coming temperature. The *ESS* criterion is a measure of the current diversity and approximation accuracy of the particle system. Over time, as the SMC algorithm generates samples from one bridge density to the next, the effective sample size typically decreases as some particles may have degenerating weights. For example, some draws of the prior may be located in areas of the parameter space that hold little weight in the intermediate density $\pi_n(\theta)$. As a result, without resampling, one expects the ESS to decrease over the iterations. Jasra et al. (2010) propose to control the decay of the ESS based on some user-chosen rate. To implement this, the ESS is calculated as:

$$ESS_n(\phi_n) = \frac{N}{\frac{1}{N}\sum_{j=1}^{N}\left(W_n^j(\phi_n)\right)^2}.$$

Crucially, the $ESS_n(\phi_n)$ in iteration n varies only by ϕ_n as previous weights and log-likelihood values are fixed in the calculation. Based on this, a decay criterion can be chosen and minimized:

$$abs(ESS_n(\phi_n) - \alpha ESS_{n-1}) = 0.$$

Conceptually, the above criterion satisfies that the current effective sample size in iteration n does not decay too much or too little from ESS_{n-1} as governed by α. Appropriate choices of α can ensure a gradual, consistent and plannable decay of the ESS. To ensure that the ESS does not decay to zero and accurate approximations are maintained, the particle system is resampled using systematic resampling whenever the ESS is less than half of the total sample size, N. As a result, the ESS goes through a repeated pattern of decay governed by α, which is followed by upward jumps close to the total sample size. The path of ϕ_n going from zero to one arguably depends on the complexity of the density. Within a given decay phase, the path is typically linear. To conclude, the crucial advantage of Jasra et al. (2010) is that the tempering schedule is neither too fast nor too slow and avoids manual calibration based on expensive test runs. Algorithm 1 summarizes the adaptive tempering procedure in a quasi-code format.

$\boldsymbol{Algorithm\ 1}$: $temperature\ adaption\ g()$

1. $Input$:
   $\phi_{n-1} =\ temperature\ at\ n-1$
   $W_{n-1}^{i} = normalized\ weights\ at\ n-1$
   $p\left(Y\middle|\theta_{n-1}^{i}\right) = loglikelihood\ values$
   $p\left(\theta_{n-1}^{j}\right) = prior\ likelihood$
   $\mu_1\left(\theta_{n-1}^{j}\right) = proposal\ density\ values$
   $ESS_{n-1} = Effective\ sample\ size\ at\ n-1$
   $\alpha = parameter\ controlling\ particle\ degeneracy$
2. $Initialization$:
   $define\ w_n^i =\ (p\left(y|\theta_{n-1}^{j}\right)p\left(\theta_{n-1}^{j}\right))^{\phi_n-\phi_{n-1}}\mu_1(\theta_{n-1}^{j})^{\phi_{n-1}-\phi_n}\ and\ W_n^i = \frac{W_{n-1}^i w_n^i}{\sum_{i=1}^{N} W_{n-1}^i w_n^i}$
   $define\ ESS_n(\phi_n) = \frac{N}{\frac{1}{N}\sum_{i=1}^{N}\left(W_n^i\right)^2}$
3. $quasi\ code$:
   $if\ ESS_n(1) \geq aESS_{n-1}$
   $\quad \phi_n = 1$
   $else$
   $\quad solve\ abs(ESS_n(\phi_n) - aESS_{n-1}) = 0$
   end

Algorithm 1: Quasi Code for the Jasra et al. (2010) adaptive tempering strategy applied to SMC sampling.

The last component of the SMC sampler is the mutation step. While the prior draws might offer sufficient coverage over the posterior, it is beneficial to adapt the particle system to the current density, $\pi_n(\theta)$, to reintroduce variation that is lost during the resampling steps. Following Herbst and Schorfheide (2016), I implement a blocked Metropolis-Hastings sampler using a mixture proposal density. The Metropolis-Hastings algorithm is a particularly important choice because the sampler leaves the particles invariant. However, Metropolis-Hastings samplers do not scale well with increasing parameter numbers. As the number of parameters increases, the rate of exploration through the posterior decreases (e.g., see Neal (2012)). If one relies on blocking and mutates a sub-vector of θ, this implies considerably higher acceptance rates. The blocked mixture proposals, $v_{n,b}^{j}$, for block, b, of particle j in iteration n of the SMC sampler then comes from the following density:

$$v_{n,b}^{j}|(\theta_{n,b}^{j}, \bar{\theta}_{n,b}, \Sigma_{n,b}) \sim wN(\theta_{n,b}^{j}, c^2\Sigma_{n,b}) + \frac{1-w}{2}N(\theta_{n,b}^{j}, c^2 diag(\Sigma_{n,b}))$$
$$+\frac{1-w}{2}N\left(\bar{\theta}_{n,b}, c^2\Sigma_{n,b}\right),$$

where $\Sigma_{n,b}$ is the particle approximation of the covariance matrix of $\theta_{n,b}$, $\bar{\theta}_{n,b}$ is the mean of the sub-vector, and c is the scaling factor of the proposal. The scaling factor is chosen based on a targeting function of the Herbst and Schorfheide (2016) design. This density has three mixture components. Firstly, it offers a standard random walk proposal $N(\theta^j_{n,b}, c^2\Sigma_{n,b})$ using the full covariance matrix with probability w. Secondly, it has a further random walk proposal $N(\theta^j_{n,b}, c^2 diag(\Sigma_{n,b}))$ with probability $\frac{1-w}{2}$, where correlations between parameters are ignored. Lastly, it features an independent proposal $N(\bar{\theta}_{n,b}, c^2\Sigma_{n,b})$ with $p = \frac{1-w}{2}$ where samples are generated at the mean. In practice, for DSGE models, it is typically the case that parameters in θ are constrained. In this case, a normal approximation as above working with θ may generate proposals out of bounds that will be rejected immediately. Here, I implement a strategy based on Amisano and Tristani (2010) based on working on a transformed parameter vector in an unconstrained space. For gamma-distributed parameters, the *log* transformation is applied. For Beta distributed parameters, an inverse sigmoid transformation is selected, and normal parameters are not transformed. To ensure that the sampler still works, the acceptance step of the Metropolis-Hastings algorithm is adjusted using the determinant of the Jacobian of the transformation.

The above described algorithm requires a number of tuning parameters in order to run. The number of particles, N, is set to 3000 and the particle degeneration parameter, α, is set to 0.9. For the mixture distribution the number of blocks is set to 5, the mixture weight, w, is equal to 0.9 and the initial scale parameter, c, is set to 0.5.

For the code implementation, I rely heavily on the initial implementation of their SMC published by Herbst and Schorfheide (2016). However, I adjust the code to include the individual changes listed above. Algorithm 2 provides a summary of the SMC strategy employed in this chapter.

Algorithm 2: *SMC sampler with blocked MCMC steps*

1. *Inputs*:
   N = number of particles
   g = function to choose next temperature
   K = Markov Kernel for MCMC steps
   N_b = number of blocks
   f = function to choose next step size
2. *Initialization*:
   $set\ n = 1\ and\ \phi_1 = 0$
   $sample\ \theta_1^i \sim \pi_1(\theta) = \mu_1(\theta)$
   $set\ W_1^i = 1\ \forall\ i \in \{1, \dots, N\}$
3. *Iteration*:
   $while\ \phi_n < 1$
   $set\ n = n + 1\ and\ choose\ \phi_n = g()$

   correction step:
   $W_n^i = \frac{W_{n-1}^i w_n^i}{\sum_{i=1}^N W_{n-1}^i w_n^i}\ \ and\ \ w_n^i = (p(y|\theta_{n-1}^j)p(\theta_{n-1}^j))^{\phi_n - \phi_{n-1}} \mu_1(\theta_{n-1}^j)^{\phi_{n-1} - \phi_n}\ \forall\ i$

   selection step:
   $ESS_n = \frac{N}{\frac{1}{N}\sum_{i=1}^N (\tilde{W}_n^i)^2}$
   $if\ ESS_n \geq 0.5 \times N$
   $W_n^i = W_n^i\ \ and\ \theta_n^i = \theta_{n-1}^i\ \forall\ i \in \{1, \dots, N\}$
   $else$
   $use\ systematic\ resampling\ with\ \{\theta_{n-1}^i, W_n^i\}_{i=1}^N$
   $generate\ sample\ \{\theta_n^i\}_{i=1}^N\ with\ W_n^i = 1$
   end

   mutation step:
   $choose\ c_n = f(c_{n-1}, acceptance\ rate)$
   $move\ particle\ \theta_n^i \sim K()\ using\ blocked\ MH\ with\ N_b$
   end

Algorithm 2: Summary of the SMC algorithm

2.4.3 Data

The following section gives a detailed description of the construction of the observable variables. For the likelihood construction, the model uses seven observable variables over a sample from 1984Q1 to 2021Q4. The sample purposefully excludes the early 1980s to avoid including periods of output volatility, similar to Sims and Wolff (2018a). The end date was the latest date for which the data set could be fully constructed. The data includes measurements for the following variables: federal consumption tax rate, τ_t^C, federal labour taxation rate, τ_t^l, federal government consumption, G_t, federal government debt, B_t, GDP, Y_t, Inflation, π_t and interest rates, i_t. Out of convenience, I drop the index t below, but variables still refer to their measurements in period t.

Starting with the fiscal variables, this chapter heavily orientates itself on the design of taxation rates developed in Jones (2002) and later used in Leeper, Plante and Traum (2010). To construct τ^c and τ^l, a few intermediate steps are needed, and all data is taken from the Bureau of Economic Analysis (BEA) and the Federal Reserve Economic Data database (FRED). To construct the consumption tax rate, overall and local consumption tax revenues and, in addition, the level of consumption are needed. Firstly, consumption tax revenues, T^c, are the taxes on production and imports. T^c Includes both excise taxes and customs duties. Consumption, C, is defined as the sum of personal consumption expenditures on nondurable goods and services.

Based on T^c, T_s^c and C, the marginal consumption tax rate can be constructed as follows:

$$\tau^c = \frac{T^c}{C - T^c \ - T_s^c}.$$

To construct the average labour income tax rate, one first needs to construct the average personal income tax rate following the methodology presented in Jones (2002).

$$\tau^p = \frac{IT}{W + \frac{PRI}{2} + CI}.$$

Capital income, CI, is defined as the sum of rental income, corporate profits and interest. The variable W represents wage and salary accruals, and PRI corresponds to the proprietor's income. Based on τ^p, one can construct the average labour income tax rate as:

$$\tau^l = \frac{\tau^p(W + PRI/2) + CSI}{EC + PRI/2},$$

where CSI are contributions to government social insurance, and EC are compensations to employees. Within the model, the tax rates enter under the following transformation: $\frac{\tau_t}{1+\tau_t}$. In practical terms, this ensures that the rate is contained in the closed interval from zero to one. To match the data to the model variables, the opposite transformation is applied to the tax rate observations.

$$\tau_t^{obs} = \frac{\tau_t}{1 - \tau_t}.$$

This means that the corresponding model variables can be directly matched as opposed to matching a non-linear transformation of the model variables, $\frac{\tau_t}{1+\tau_t}$, to the data. In practical terms, the difference between τ_t and $\frac{\tau_t}{1-\tau_t}$ tends to be fairly small away from the boundary.

Government consumption, G, is set as the sum of federal government consumption and government net purchases of non-produced assets minus government consumption of fixed capital. Government Debt, B, is collected as market value US federal debt from the Dallas federal reserved database.

Departing from the government side, the data construction requires three more series. Firstly, GDP is collected as seasonally adjusted gross domestic product from the FRED database. Secondly, inflation is constructed using the implicit price deflator of GDP, and thirdly, interest rates are defined as 3-Month market rates of treasury bills. Both series were collected from the FRED database as well. To construct model variables, government consumption, G_t, gross domestic product, Y_t, and government debt, B_t, are first deflated using the GDP deflator to obtain real-valued variables.

In a further step, some variables need to be detrended. Macroeconomic variables often include trends, and it is important to account for this in either the data construction or the modelling. In practice, the choice of procedure determines some of the features of the data and, as such, needs to be carefully evaluated. Pfeifer (2018) explores the current and popular approaches to account for trend mechanics in DSGE models. One approach comes in the form of including a growth rate transformation of trending variables. For a lot of data sets like GDP data, growth rates are typically approximately stable across time. This approach is, for example, followed in the canonical paper by Smets and Wouters (2007). An alternative approach is to linearly detrend the log of the nominal variables. The resulting data can be interpreted as steady state deviations from a trend and is used in Leeper, Plante and Traum (2010). The detrending options are not limited to these two but also include the Hodrick-Prescot filter, directly modelling the trend and others.

In relation to this, Pfeifer (2018) argues two things. Firstly, detrending data is meant to filter out business cycle mechanics that are not reflected in the model and preserve features that are. Therefore, the various detrending options remove related features in the data that are arguably generated by the business cycle. However, the removed features may vary in design across the detrending strategies. Secondly, Pfeifer (2018) argues that the difference in detrending strategies mostly reflects a priori preferences of the economist designing the model. In particular, it relates to which features of the data are assumed to be related to the business cycle and which are not. In this application, I utilized linear detrending of the log variables. This approach presents several advantages. Firstly, the model of fiscal policy is closely based on Leeper, Plante and Traum (2010). For comparison purposes, it seems advantageous to include the data in a similar fashion. Secondly, linear detrending typically leads to deeper business cycles in comparison to growth rates. This is consistent with the prior belief that fiscal policy goes through long and persistent policy cycles, and similarly, output business cycles are assumed to be fairly deep. The last advantage is computational. Linear detrending produces variables that have the interpretation of percentage steady state deviations. These variables are directly measured by DSGE models. In comparison, growth rates are typically not directly measured and can be constructed by subtracting steady deviations from two periods. For the simulation, the latter approach requires keeping track of redundant variables and can increase simulation time.

Based on this, all data series but inflation and interest data are detrended using a linear trend on the log values. The resulting interpretation of the model variables is as log deviations from their steady state. For inflation and the interest rate, it is assumed that the data comes from a stationary, non-trending distribution. Further, inflation and interest rates are transformed using a log transformation.

2.4.4 Code implementation

Estimating higher-order DSGE models can be very time-consuming, computationally complex and can require a lot of code development as in-built libraries may not always be suitable for all individual projects. Classical inference strategies with sequential sampling steps, like the random walk Metropolis-Hastings algorithm, can require estimation time in the magnitude of weeks to months depending on the individual researchers computational set up and number of sampling steps. Together, this makes these types of projects, which are already difficult to implement, prohibitive from a time cost perspective unless the researcher has access to state-of-the-art computing systems. To make the estimation of chapter 2 feasible, I attempt to improve on the standard estimation techniques by combining different empirical and technical approaches.

A feature of modern computational developments is that CPUs or GPUs are not getting much faster on a per core basis, but parallel computations are where most improvements are being made. In order to exploit this, I focus on the use of Sequential Monte Carlo as proposed by Herbst and Schorfheide (2016) for the use in economics. I propose a specific Sequential Monte Carlo sampler heavily inspired by Buchholz, Chopin and Jacob (2021) and Jasra et al. (2010). The sampler attempts to make use of the computationally efficient parallelism of Sequential Monte Carlo sampler, while avoiding spending effort into exploring irrelevant areas of the posterior or already well explored tempered distributions.

Parallelism aside, another crucial factor is the time per likelihood evaluations. Unlike Kalman filter based likelihood evaluations, particle filters rely on large array operations and are quite time consuming. To improve the computation time, I pass the main time consuming, large array operations to a GPU similar to neural network application for gradient backpropagation. Lastly, I attempt to make use of good coding practices by focusing on utilizing the LAPACK libraries whenever possible or utilizing MATLABs symbolic toolbox for the model creation.

Together, the estimation time is cut down to around 5 days, which implies a reduction of up to 94% depending on the chosen comparison basis. The main influences in this process are provided by Gomme and Klein (2011), Schmitt-Grohe and Uribe (2004), Herbst and Schorfheide (2016), Buchholz, Chopin and Jacob (2021), Jasra et al. (2010) and neural network applications.

I begin the code development with a replication exercise on the Amisano and Tristani (2010) model on its original data set. During the replication process, I was greatly aided by Amisano and Tristani (2010) sharing their code with me, which allowed me to double-check my work and improve on it. The main estimation in this paper features more model states than the original model, which increases the estimation time quite drastically. To combat this, a substantial part of the work went into finding strategies to reduce the estimation time, and this chapter utilizes three main improvements: adaptive Sequential Monte Carlo estimation, parallelization and GPU use for large array operations. The three approaches aim to reduce computation time by reducing the required number of likelihood evaluations to a minimum, evaluating likelihoods in parallel and optimizing speed per likelihood evaluation.

The first two improvements go hand in hand. Traditionally, the Metropolis-Hastings algorithm is employed to estimate DSGE models. The algorithm is very useful, and the implementation is straightforward. However, the estimation time tends to be large as it can require a large number of likelihood evaluations to explore the posterior, and the algorithm works in a sequential fashion. That means it scales roughly linearly to the number likelihood evaluation and estimation time is approximately equal to $n_{eval} * t_{likelihood\ eval.}$.

SMC algorithms, as proposed by Herbst and Schorfheide (2016) for the use in economics, can be used to bring down the estimation time. Firstly, the adaptive SMC algorithm constructed in this chapter implicitly chooses the number of likelihood evaluations required for the estimation based on the adaptive tempering schedule proposed in Jasra et al. (2010). Based on my testing, the overall number of likelihood evaluations tends to be lower than in typical Metropolis-Hastings applications in the literature, and it avoids expensive test runs. Secondly, unlike the Metropolis-Hastings algorithm, the SMC algorithm can be run in parallel, which brings the estimation time down to $\frac{n_{eval} \times t_{likelihood\ eval}}{n_{cores}}$.[14]

The last main point of improvement focuses on reducing the time per likelihood evaluation. Unlike linear estimations using the Kalman filter, non-linear estimations using particle filter methods require large array operations due to the second-order terms. Large array operations are

[14] That figure is approximate and ignores communication overheads, and other factors. However, in most applications the gain is still substantial.

most conveniently and efficiently done on GPUs, and this strategy is also frequently employed in other fields like neural network estimations. Based on the GPU setup, I reduce the likelihood evaluation time by 58%.

While the previous points delivered the most significant increases in performance, the following aspects help to improve code performance further: focus on writing code optimized for the inbuilt LAPACK libraries, symbolic differentiation of model files, faster model solution strategies and others. Overall, the techniques described above and, in the appendix, reduce the estimation from weeks to days. Additionally, the estimation time can be reduced by up to 94% depending on the selected comparison basis.

For a more detailed description of the individual improvement strategies, see the appendix. I also provide re-estimation results for the Amisano and Tristani (2010) model. That includes estimations of the Amisano and Tristani (2010) model in its linear and non-linear form using the Metropolis-Hastings approach employed by Amisano and Tristani (2010) and an SMC estimation of the non-linear model version.

2.4.5 Posterior estimates

The following section presents posterior estimates for the model parameters. The results are summarised in Table 2.3, Table 2.4 and Table 2.5.

To start off, the habit formation parameter is estimated to be around 0.45, which is a bit lower than typical estimates of around 0.7. However, it is fairly close to the Leeper, Plante and Traum (2010) estimate of 0.5. Price indexation, l, is estimated to be quite a bit higher than in the Amisano and Tristani (2010) model at around 0.58 and much closer to estimates in Sims and Wolff (2018a). The Calvo pricing parameter is similar to Sims and Wolff (2018a) and Smets and Wouters (2007). The Taylor rule features a strong response to inflation deviations from the target as governed by ψ_π with a posterior mean of 1.9. Output growth responses are smaller by comparison. The autoregressive parameters all show estimates with fairly high persistence over 0.9. The autoregressive parameter for productivity is fairly close to a unit root process. Highly persistent shocks are not unusual, though this one is particularly persistent. The origin for the high estimate seems to be in the original Amisano and Tristani (2010) paper, who estimate p_a identically in their non-linear estimation.

Table 2.3: Posterior estimates for core model parameters

para	mean	sd.	para	mean	sd.
β	0.99611	0.00091	p_π	0.96910	0.02319
$\gamma - 1$	2.88668	1.24617	σ_{τ^l}	0.01904	0.00241
h	0.45076	0.15845	σ_{τ^c}	0.03001	0.00383
$\phi - 1$	0.94969	0.36675	σ_T	0.05297	0.01142
$\theta - 1$	7.46023	3.58253	σ_G	0.02339	0.00220
ζ	0.72244	0.08651	σ_a	0.02533	0.00492
l	0.57743	0.11304	σ_i	0.00143	0.00034
$\psi_\pi - 1$	0.90019	0.22510	σ_π	0.00156	0.00028
ψ_y	0.12943	0.04407	τ^l	0.23014	0.00113
p_{τ^l}	0.95814	0.02456	τ^c	0.01556	0.00139
p_{τ^c}	0.97063	0.02592	s_g	0.05984	0.00117
p_Z	0.93096	0.03765	s_b	0.49469	0.01114
p_G	0.95531	0.02423	π	0.00564	0.00027
p_a	0.99948	0.00123			
p_i	0.91657	0.02164			

The linear fiscal parameters show quite standard posterior estimates for the debt and output responses. The labour tax debt coefficient, $\mu_{\tau^l,B}$, is estimated to be somewhat higher than in Leeper, Plante and Traum (2010) at 0.21, while in turn, the transfer debt coefficient, $\mu_{Z,B}$, is estimated to be lower at 0.09. The parameters that govern productivity and inflation responses overall do not deviate far from zero relative to the prior standard deviation. But specific responses like transfers to productivity, $\mu_{Z,A}$, estimated at 0.04 and government consumption to inflation, $\mu_{G,\pi}$, estimated at 0.02, may prove to improve dynamics.

Table 2.4: Posterior estimates for linear fiscal parameters

para	mean	sd.
$\mu_{\tau^l,Y}$	0.18121	0.12679
$\mu_{\tau^l,B}$	0.21419	0.07318
$\mu_{\tau^l,A}$	0.00492	0.11651
$\mu_{\tau^l,\pi}$	0.00706	0.16307
$\mu_{Z,Y}$	0.20610	0.14094
$\mu_{Z,B}$	0.08613	0.09510
$\mu_{Z,A}$	0.04356	0.10380
$\mu_{Z,\pi}$	-0.01633	0.14512
$\mu_{G,Y}$	0.05127	0.08664
$\mu_{G,B}$	0.34541	0.12567
$\mu_{G,A}$	-0.01468	0.09610
$\mu_{G,\pi}$	0.02190	0.10908

Moving on to the non-linear fiscal parameters, about a third of the parameters deviate from the prior mean by at least around half a standard deviation. However, that also includes some parameters that deviate quite substantially. For labour taxation, the interaction terms between output and inflation, $\varphi_{\tau^l,Y,\pi}$, and debt with itself, $\varphi_{\tau^l,B,B}$, are particularly pronounced in their deviation from the prior with a posterior mean of 0.19 and 0.48, respectively. For transfers, the interactions between output and productivity seem to be of particular importance, with a posterior mean of -0.32. For government consumption, there are a few parameters of moderate importance: $\varphi_{G,Y,Y}$, $\varphi_{G,Y,B}$, $\varphi_{G,Y,\pi}$ and $\varphi_{G,B,A}$. All of these deviate from zero by about half a prior standard deviation. Based on this, it seems to be the case that the data provides evidence in favour of non-linear fiscal rules, and, specifically, it shows that standard linear rules can be improved upon by capturing business cycle dependency. To explore the overall influence of the non-linear fiscal parameters on the gradients of the fiscal response functions, section 2.5.5 traces out the gradients across time to interpret the parameters in a joint fashion.

Table 2.5: Posterior estimates for non-linear interaction parameters

para	mean	sd.	para	mean	sd.
$\varphi_{\tau^l,Y,Y}$	-0.00672	0.22839	$\varphi_{Z,A,A}$	0.13796	0.22957
$\varphi_{\tau^l,Y,B}$	0.06161	0.23894	$\varphi_{Z,A,\pi}$	-0.08409	0.26127
$\varphi_{\tau^l,Y,A}$	-0.02342	0.30735	$\varphi_{Z,\pi,\pi}$	-0.05243	0.29616
$\varphi_{\tau^l,Y,\pi}$	0.19194	0.30617	$\varphi_{G,Y,Y}$	0.11031	0.25147
$\varphi_{\tau^l,B,B}$	0.47587	0.10874	$\varphi_{G,Y,B}$	0.14026	0.32830
$\varphi_{\tau^l,B,A}$	-0.07431	0.30192	$\varphi_{G,Y,A}$	-0.04014	0.23953
$\varphi_{\tau^l,B,\pi}$	0.01719	0.16656	$\varphi_{G,Y,\pi}$	0.10777	0.30804
$\varphi_{\tau^l,A,A}$	-0.01476	0.20969	$\varphi_{G,B,B}$	0.04411	0.25756
$\varphi_{\tau^l,A,\pi}$	0.03715	0.22585	$\varphi_{G,B,A}$	-0.13786	0.23030
$\varphi_{\tau^l,\pi,\pi}$	-0.05738	0.20750	$\varphi_{G,B,\pi}$	-0.01568	0.21275
$\varphi_{Z,Y,Y}$	-0.07245	0.27824	$\varphi_{G,A,A}$	0.07878	0.29301
$\varphi_{Z,Y,B}$	-0.08677	0.20179	$\varphi_{G,A,\pi}$	-0.01862	0.25692
$\varphi_{Z,Y,A}$	-0.32708	0.29289	$\varphi_{G,\pi,\pi}$	-0.00060	0.21107
$\varphi_{Z,Y,\pi}$	0.08429	0.31447	$\psi_{\pi,\pi}$	-0.61796	1.11443
$\varphi_{Z,B,B}$	-0.09492	0.12617	$\psi_{\pi,Y}$	1.20370	1.29728
$\varphi_{Z,B,A}$	-0.03489	0.28670	$\psi_{Y,Y}$	-0.63064	1.71219
$\varphi_{Z,B,\pi}$	0.04353	0.27273			

The last category of non-linear parameters is the non-linear parameters in the interest rate rule. $\psi_{\pi,\pi}$, $\psi_{\pi,Y}$ and $\psi_{Y,Y}$ all deviate more than half a prior standard deviation from zero. $\psi_{\pi,\pi}$ is estimated to be negative at -0.61796. That means that the responsiveness to inflation deviations from the target, $(\pi_t - \pi_t^*)$, is decreasing in $(\pi_t - \pi_t^*)$. In practice, that paints a picture of a government that always increases the interest rate in response to inflationary pressure. But, as the pressure keeps building up, interest responses become smaller and smaller, indicating limits to interest rate policy. The same holds for the responsiveness of interest rates to output growth governed by $\psi_{Y,Y}$. $\psi_{Y,Y}$ is estimated to be negative at -0.63064. As output growth increases, the gradient of the interest rate with respect to output growth decreases. Curiously, the interaction term $\psi_{\pi,Y}$ is estimated at 1.20370, meaning that the interaction effects are positive.

The appendix presents estimation diagnostics.

2.5 Fiscal policy effectiveness

2.5.1 Mechanics of state dependency

In the following, I explore the dynamics of impulse responses in second-order pruned systems and show how they relate to the business cycle. To start off, I establish a comparison basis using the first-order DSGE approximation. The canonical linear system is defined as follows:

$$x_t^L = H_x x_{t-1}^L + \sigma J v_t, \qquad v_t \sim N(0, I),$$

where x_t^L is a $(n_x \times 1)$ vector of model states and H_x is a $(n_x \times n_x)$ matrices that governs the system dynamics. v_t is the structural shock vector of size $(n_v \times 1)$, which is normally distributed with mean zero and an identity covariance matrix. The impact of v_t on x_t^L is governed by σ and J where σ is a perturbation scalar typically set to one and J is a $(n_x \times n_v)$ matrix. If a shock v_t occurs today, then it has an immediate impact on x_t^L today, but the impact also transitions through the system as governed by H_x and can influence future linear states. To be precise, the expectation of x_{t+h}^L can be constructed as:

$$E(x_{t+h}^L | x_{t-1}^L, v_t) = H_x^{h+1} x_{t-1}^L + H_x^h J v_t.$$

The expectation of x_{t+h}^L depends linearly on both the initial conditions of the economy, x_{t-1}^L, and the shock, v_t, to which the economy is subjected. To construct the impulse response in the linear system, one compares a world in which the shock happened to one where It did not:

$$IRF_{t+h} = E(x_{t+h}^L | x_{t-1}^L, v_t) - E(x_{t+h}^L | x_{t-1}^L, v_t = 0) = H_x^h J v_t$$

The above equation can be interpreted as the difference between the expected state of the economy at time h when the shock happened and the economy where it did not. For linear models, the initial conditions of the economy cancel out, and the only important factors are v_t and the horizon h. Consequently, the difference between the two paths is independent of when the shock occurs. For this reason, it is common practice in economics to view impulse responses as conducted at the steady state of the economy. At the steady state, $H_x^{h+1} x_{t-1}^L$ is always equal to zero, and one only has to construct $H_x^h J v_t$. Furthermore, this approach can improve the interpretation as the impulse response can be viewed as the steady state deviation of the model states caused by the shock. However, it also retains the difference between the expected paths for some initial state x^L_{t-1}.

For non-linear models, the situation becomes more complicated, and I rely on the pruned second-order approximation as in Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017) or Amisano and Tristani (2010). Using the pruned system has several advantages. Firstly, higher-order approximations are almost always pruned as the pruning can ease simulation problems and preserves a lot of the original dynamics. Secondly, it simplifies the impulse responses analysis and pinning down the relationship to the business cycle becomes easier. The main change in the pruned second-order system is that the impact of the second-order terms on the quadratic states, x^Q_t, is governed by an auxiliary linear system:

$$x^L_t = H_x x^L_{t-1} + \sigma J v_t, \qquad v_t \sim N(0, I),$$

$$x^Q_t = 0.5 * h_{\sigma\sigma} + H_x x^Q_{t-1} + 0.5 * H_{xx}(x^L_{t-1} \bigotimes x^L_{t-1}) + \sigma J v_t, \qquad v_t \sim N(0, I).$$

where $h_{\sigma\sigma}$ is a vector of dimension $(n_x \times 1)$ and H_{xx} is a $(n_x \times n_x^2)$ matrix. The pruned second-order system can be rewritten into a linear system using an augmented state vector, $z_t = [x^L_t{}', x^Q_t{}', (x^L_t \bigotimes x^L_t)']'$ and shock vector, ζ_t:

$$z_t = c_2 + A_2 z_{t-1} + B_2 \zeta_t.$$

Here, c_2 is a $((2n_x + n_x^2) \times 1)$ constant vector, A_2 is a matrix of size $((2n_x + n_x^2) \times (2n_x + n_x^2))$ and B_2 is of size $((2n_x + n_x^2) \times (n_v + n_v^2 + 2 * n_x n_v))$. For the exact design of the matrices, see Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017) or appendix. The shock vector ζ_t is designed as follows:

$$\zeta_t = \begin{bmatrix} v_t \\ (v_t \bigotimes v_t) - vec(I_{n_v}) \\ (v_t \bigotimes x^L_{t-1}) \\ (x^L_{t-1} \bigotimes v_t) \end{bmatrix} = \begin{bmatrix} v_t \\ (v_t \bigotimes v_t) - vec(I_{n_v}) \\ (v_t \bigotimes x^L_{t-1}) \\ P(v_t \bigotimes x^L_{t-1})Q \end{bmatrix},$$

where P and Q are permutation matrices. As the new augmented second-order system is linear, the impulse response can be constructed identically as:

$$IRF_{t+h} = E_t(z_{t+h} | z_{t-1}, \zeta_t) - E_t(z_{t+h} | z_{t-1}, \zeta_t = 0) = A_2^h B_2 \zeta_t.$$

The main difference here is that unlike in the first-order system, the shock vector, ζ_t, depends on the initial linear conditions. The key feature of the second-order system is that it matters for the effects of policy interventions when the shock is conducted. This is, for example, what Sims and Wolff (2018a) exploit in their paper. For a given v_t and forecasting horizon, the impulse response of state i is linear in the linear initial conditions, x^L_{t-1}.

Consider a state i, then the impulse response as a function of the linear states is defined as:

$$IRF^i_{t+h}(x^L_{t-1}) = \gamma A_2^h B_2 \zeta_t,$$

where γ is a $(1 \times (2n_x + n_x^2))$ row vector with all elements equal to zero but entry i. $IRF^i_{t+h}(x^L_{t-1})$ is affine in x^L_{t-1} if and only if it is both convex and concave. Take two initial condition vectors x, $y \in \mathbb{R}^{n_x}$ and $\lambda \in [0,1]$, then $IRF^i_{t+h}(x)$ is affine if and only if the following holds:

$$IRF^i_{t+h}(\lambda x + (1-\lambda)y) = \lambda IRF^i_{t+h}(x) + (1-\lambda) IRF^i_{t+h}(y).$$

The above may be rewritten as:

$$\gamma A_2^h B_2 \big(\zeta_t(\lambda x + (1-\lambda)y) - \lambda \zeta_t(x) - (1-\lambda)\zeta_t(y)\big) = 0.$$

Disregarding the trivial case of $\gamma A_2^h B_2 = 0$ and making use of the additive properties of the Kronecker product, $\zeta_t(\lambda x + (1-\lambda)y)$ may be rewritten as:

$$\begin{aligned}
\zeta_t(\lambda x + (1-\lambda)y) &= \begin{bmatrix} v_t \\ (v_t \otimes v_t) - vec(I_{n_v}) \\ (v_t \otimes (\lambda x + (1-\lambda)y)) \\ P(v_t \otimes (\lambda x + (1-\lambda)y))Q \end{bmatrix} = \begin{bmatrix} v_t \\ (v_t \otimes v_t) - vec(I_{n_v}) \\ (v_t \otimes (\lambda x)) + (v_t \otimes ((1-\lambda)y)) \\ P((v_t \otimes (\lambda x)) + (v_t \otimes ((1-\lambda)y)))Q \end{bmatrix} \\
&= \lambda \begin{bmatrix} v_t \\ (v_t \otimes v_t) - vec(I_{n_v}) \\ (v_t \otimes x) \\ P(v_t \otimes x)Q \end{bmatrix} + (1-\lambda) \begin{bmatrix} v_t \\ (v_t \otimes v_t) - vec(I_{n_v}) \\ (v_t \otimes y) \\ P(v_t \otimes y)Q \end{bmatrix} \\
&= \lambda \zeta_t(x) + (1-\lambda)\zeta_t(y).
\end{aligned}$$

Thus, this leads us to conclude that IRF^i_{t+h} is affine in the initial, linear conditions. This has an important consequence for the design of impulse responses. To illustrate this point, suppose that we are interested in the effects of government consumption shocks on output and assume it is known that the effects are negatively related to the interest rate (i.e., government shocks are more effective in low interest rate periods). If the IRF^i_{t+h} is of the above design, then mechanically, it is feasible for the IRF^i_{t+h} of output to government consumption shocks to reverse sign for sufficiently high interest rates. When and at what point this happens depends on the slope of IRF^i_{t+h} with respect to the interest rate. However, while theoretically, this can happen, it does not mean it is likely from an empirical view. In a reasonable setting, interest rates may never get sufficiently large to reverse the sign in the impulse responses.

Later in the chapter, I use the fact the IRFs are affine transformations of the initial conditions to run linear regressions. In particular, I regress $IRF^i_{t+h}(x^L_{t-1})$ on x^L_{t-1} using sampled states to pin down the exact relationship. Using this technique, one can ask precise questions like "How is the impact of structural shocks on the economy governed by the initial conditions of that economy?" and "When do specific structural shocks become more or less effective at stimulating the economy?". A question that remains is what happens if the shock is changed. Unfortunately, the interaction terms in $(v_t \bigotimes v_t)$ create fundamental problems and IRF^i_{t+h} is not affine in v_t and x^L_{t-1}. Consequently, for a new shock vector, the regression strategy has to be repeated. Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017) also explore the mechanics of impulse responses but for the base system and not the augmented system. Their results confirm that shocks are not scalable because of the second-order shock terms. Further, they also show that the impulse responses depend on the linear states. I extend the analysis by explicitly proving that the way the impulse responses depend on the initial conditions is in a linear fashion.

2.5.2 Impulse response functions at and around the steady state

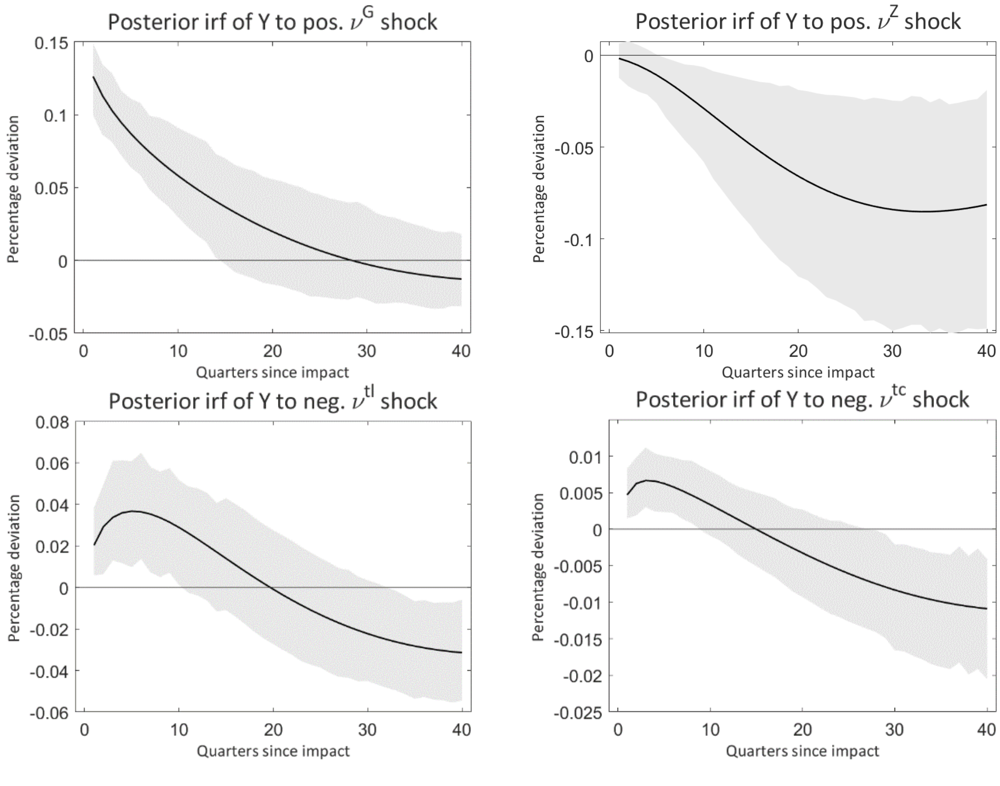
The section analyses the business cycle dependency of impulse responses to fiscal policy shocks. First, I look at the impulse responses of output in the linear states of the DSGE model, and I focus on labour taxation, consumption taxation, government consumption and transfer shocks. As a second step, I explore the impulse responses of output as a non-linear state but when fiscal policy is conducted at the steady state. In the third and last step, I conduct impulse response analysis of output as a non-linear variable over the business cycle by sampling states from the unconditional distribution. The goal of this section is to build intuition on which factors are of importance in conducting impulse responses in non-linear models and ascertain the relevance of the initial conditions for fiscal policy.

For impulse responses of the linear states, it does not matter when fiscal policy is constructed, as the difference in expected paths is a constant. The only crucial component is the type of shock, v_t. In this application, fiscal policy is conducted as one standard deviation shock to the fiscal instruments. For tax variables, I only consider tax cuts as the main point of interest.[15] For each impulse response, I simulate 500 paths to construct mean and highest posterior density intervals. The dynamics of the linear impulse responses are governed by the following equation:

$$IRF_{t+h} = H_x^h J v_t,$$

which only depends on v_t. Fig. 2.1 presents the impulse responses for the linear states. The linear impulse responses can be interpreted in two ways. Firstly, they may be interpreted as percentage deviations from the steady state, and secondly, they can be understood as the percentage deviation from the path where the shock did not occur.

Fig. 2.1: linear impulse responses of output to fiscal shocks



Notes: Impulses responses of linear output to one standard deviation fiscal shocks. The solid line is the mean impulse response; the grey shaded is the 95% highest posterior density interval.

[15] The reason is that the non-linear impulse responses are not linear in the shock vector, and tax increases cannot simply be rescaled to tax cuts by multiplying by minus one.

Overall, the linear impulse responses are entirely standard and compare well to, for example, the Leeper, Plante and Traum (2010) results. Government consumption increases have their largest impact on output immediately and fade afterwards. On Impact, a government consumption shock raises output by roughly 0.14% relative to the path. In the medium to long run, the effects of a government consumption shock turn negative based on the financing rules. Transfer shocks do not have a stimulative impact on output and decrease output in the long run. Both labour and consumption taxation cuts stimulate output on impact by 0.02% and 0.005% relative to the path, respectively. In typical fashion, the impact of tax cuts peaks at about two to three years, decaying afterwards. Based on the financing rules in this chapter, after about four to five years, the impact of tax cuts becomes negative and afterwards returns to the steady state. One key unifying factor between all of these is that the model predicts relatively tight highest posterior density intervals for the impulse responses. For example, the highest posterior density interval for government consumption shocks ranges from roughly 0.1% to 0.15%. In a sense, the model is highly confident in the range of effects that fiscal stimulus can have.

For impulse responses of the non-linear states to fiscal policy conducted at the steady state, the mechanics change as follows. For the quadratic states, it is useful to invoke the pruned second-order system representation of the impulse responses:

$$IRF^{i}_{t+h}(x^{L}_{t-1}) = \gamma A_2^h B_2 \zeta_t,$$

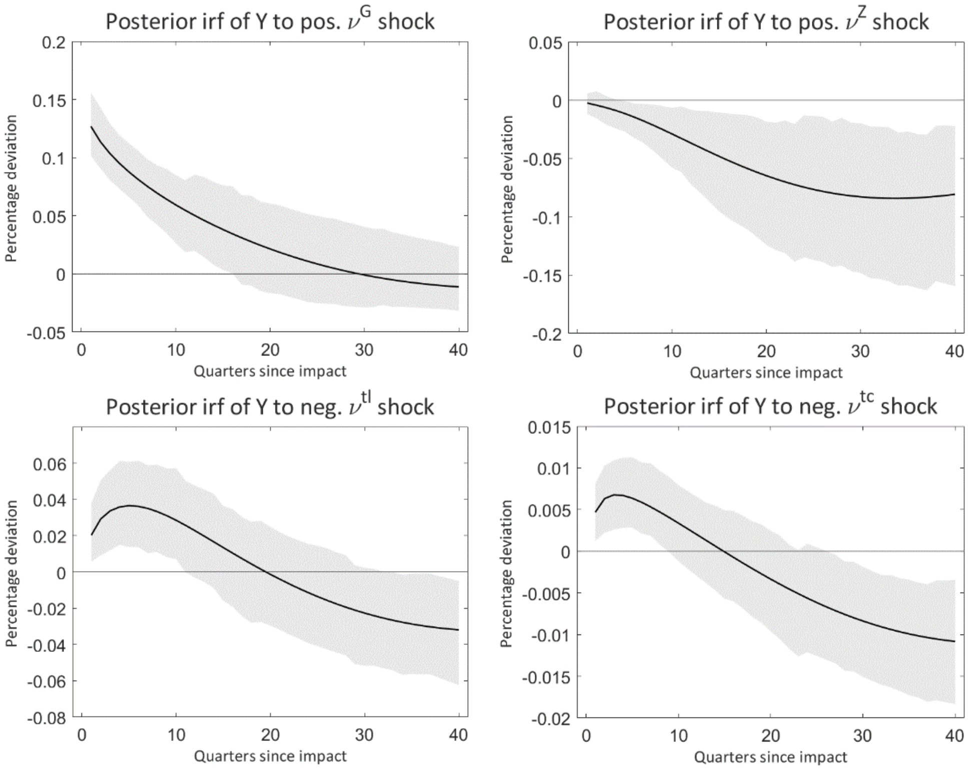
$$\zeta_t = \begin{bmatrix} v_t \\ (v_t \bigotimes v_t) - vec\left(I_{n_v}\right) \\ (v_t \bigotimes x^{L}_{t-1}) \\ (x^{L}_{t-1} \bigotimes v_t) \end{bmatrix}.$$

Fiscal policy conducted at the steady state implies that x^{L}_{t-1} is equal to $0_{(n_x \times 1)}$. The augmented shock vector reduces to the following:

$$\zeta_t|_{steady\ state} = \begin{bmatrix} v_t \\ (v_t \bigotimes v_t) - vec\left(I_{n_v}\right) \\ \left(v_t \bigotimes 0_{(n_x \times 1)}\right) \\ \left(0_{(n_x \times 1)} \bigotimes v_t\right) \end{bmatrix}.$$

These impulse responses still do not feature any state dependency on the business cycle but feature the full non-linear dynamics of the DSGE model. As can be seen in Fig. 2.2, the key result is that the impulses are, to all intents and purposes, indistinguishable from the linear impulse responses. Mechanically, this is exactly what is expected. The linear DSGE is an approximation of the non-linear set of equations that govern the full DSGE, and the equations are approximated around the steady state. At or around the steady state, linear and non-linear models will typically predict very similar dynamics. Only when the economy moves away from the steady state can the higher-order terms begin to bite. Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017) show this analytically by proving that the first-order system is second-order accurate at the steady state.

Fig. 2.2: non-linear output impulse response to fiscal shocks at steady state

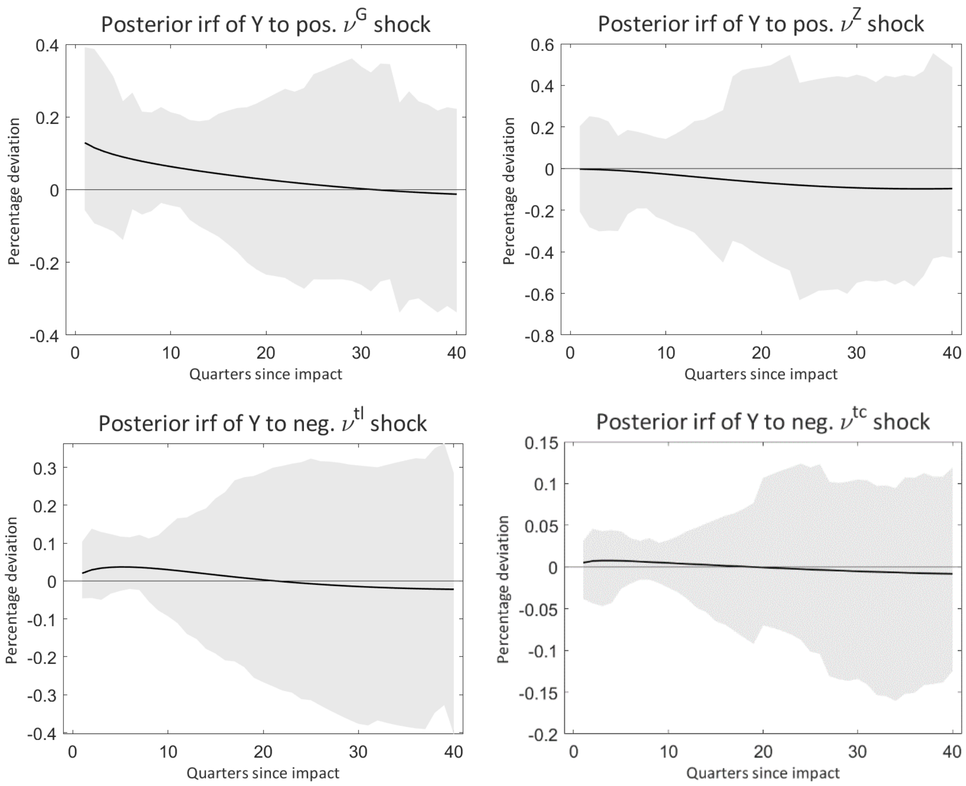


Notes: Impulses responses of non-linear output to one standard deviation fiscal shocks evaluated at the steady state. The solid line is the mean impulse response; the grey shaded is the 95% highest posterior density interval.

The last scenario for this section is based on the impulse responses of the non-linear states in response to fiscal shocks in various economic circumstances. Here, the economic circumstances are sampled from the unconditional state distribution and should generate reasonably realistic conditions as the model is fully estimated. The impulse responses are presented in Fig. 2.3. The mean response to fiscal shocks remains the same as before. However, the observable range of effects increases substantially. For example, for government consumption shocks, the mean response in linear states was about 0.15% and ranged from 0.1% to 0.20% relative to the path. Here, the highest posterior density interval ranges from below zero to about 0.4%. The upper highest posterior density bound is almost twice as large. Unlike the highly confident linear impulse responses, the non-linear impulse responses over the business cycle exude uncertainty. This result supports several factors often found to be important in impulse response analysis. Firstly, sample selection is key to the scale of effects of fiscal policy, and secondly, the timing of fiscal policy can matter significantly (i.e., fiscal policy in recessions versus at the steady state). As such, the non-linear impulse responses found here incorporate a much wider range of results found in the literature and relate them to the initial conditions of the economy.

At first glance, the negative effects of government consumption shocks are surprising and highly unusual in DSGE models. The impulse responses are affine functions and linear in the initial conditions for a given shock, as shown in the previous section. Mechanically, it is, therefore, feasible for impulse responses to reverse signs given the right economic conditions. What is happening here is that the government rules for fiscal and monetary policy are exceptionally rich, and thus, fiscal and monetary policy do not occur in isolation. If one views the governmental mechanism in unity, then it is reasonable that for a specific policy mix under certain business cycle conditions, impulse responses may deliver unusual results.

Fig. 2.3: non-linear output impulse response to fiscal shocks around the cycle



Notes: Impulses responses of non-linear output to one standard deviation fiscal shocks evaluated at randomly sampled states. The solid line is the mean impulse response; the grey shaded is the 95% highest posterior density interval.

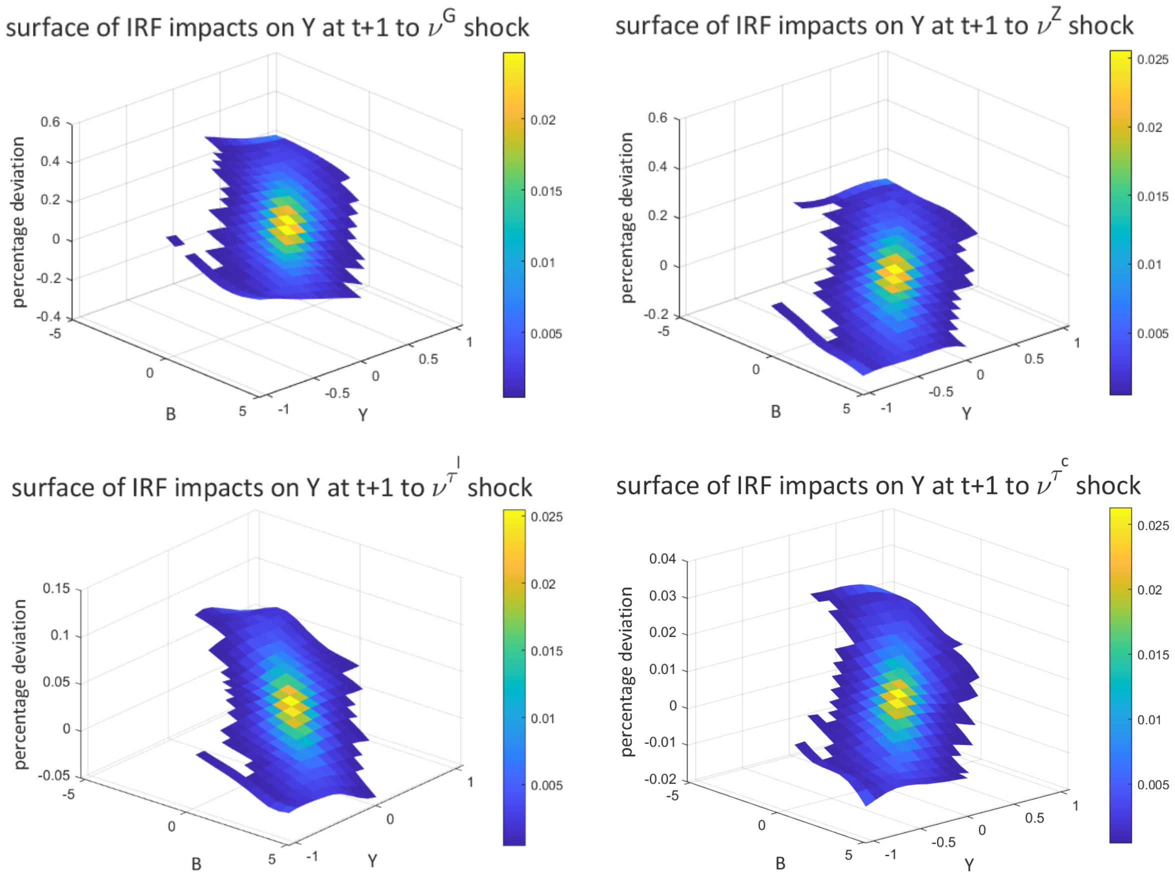
To sum up, the estimated model shows standard and relatively precise impulse responses to fiscal shocks in the linear framework. Changing to the non-linear framework but conducting fiscal policy at the steady state does not offer additional behaviour or conclusions. However, viewing impulse responses over the business cycle in the non-linear form offers additional insights. The key result is that impulse responses are much more diffuse in their effects and incorporate a much broader range of results, unlike the linear counterparts. Based on this, going forward, I will explore if it is possible to reduce some of this uncertainty by narrowing down the relationship between impulse responses and initial conditions.

2.5.3 Relationship between policy effectiveness and the initial conditions

The previous section argued that accounting for the initial conditions can substantially increase the uncertainty in the effects of fiscal policy. In this section, I further explore the relationship via two avenues. The first avenue is based on a visualization strategy using 3D plots. However, while this approach provides useful intuition, it is limited in its applicability as there are other variables that are not controlled for. The second avenue aims to formalize the results by employing a linear regression strategy to pin down the exact relationship between the impulse responses at a given horizon to the initial conditions.

To visualize the dynamics, I sample state vectors from the unconditional state distribution evaluated at sampled posterior parameter vectors. For each state vector, a fiscal intervention is conducted, and the impact of the policy intervention is recorded for the first quarter when the fiscal shocks start affecting the economy. In this case, as the type of shock matters and the impulses are not symmetric, I focus on policy interventions aimed at boosting output: tax cuts and spending increases. The size of the shock corresponds to a one standard deviation shock. The impulse responses at the given horizon are then plotted based on the initial conditions for output and debt, which are often used as the most relevant variables in the fiscal ruleset. To aid visualization, impulse responses are averaged and smoothed across a grid to create a surface, and further, highly unlikely events are omitted ($prob < 0.000\dot{3}$). To complement the surface, the graph includes colour scaling, which includes the percentage frequency of the tile in the overall remaining sample.

Fig. 2.4: 3d slices of impulse responses of output to fiscal shocks



Notes. 3D slices of impulse responses of output to fiscal stimulus shocks. For expenditure variables, a one standard deviation increase is considered, while for tax variables, a one standard deviation decrease is considered. The impulse responses are plotted over output (Y) and debt (B) as initial conditions.

Fig. 2.4 allows for several conclusions to be drawn. Firstly, just like in the previous section, the 3D graphs indicate a significantly larger variation of the effects of policy shocks in comparison to policy conducted at the steady state. For example, the slice of government consumption impulse responses, IRF^Y_{t+1}, varies from around zero to 0.4 depending on the initial conditions. As such, it matches the range of the IRF^Y_{t+1} observed in the previous section. Secondly, the impulse responses show a clear association with the initial condition for output and debt. For all fiscal interventions, it seems to be the case that scenarios of high output and low government debt are generally associated with more effective stimulus. However, this does not mean higher output and lower debt cause policy to be more effective, as other correlated and relevant variables ought to be considered.

To formalize the relationship, going forward, I utilize a linear regression approach. This is a useful approach as it allows for a quantification of the relationship because, based on section 2.5.1 , the impulse responses are linear functions in the initial conditions for a given shock. Therefore, a linear regression utilizes the correct functional form. Furthermore, unlike the graphical approach, this methodology allows me to control for all the relevant variables at the same time.

For the regression, samples are created in an analogous fashion as for the graphs. Then, the IRF_{t+1}^{Y} of output are regressed on variables in x_{t-1}^{L} for a given shock. I exclude the structural shocks for the fiscal rules as they do not offer an easily interpretable meaning. However, this will not affect the coefficient estimates for the remaining states.[16] The results are presented in Table 2.6 and Table 2.7. Further tables for IRF_{t+4}^{Y} on initial conditions are provided in the appendix.

Table 2.6: Regression of IRFs on impact of output to gov. consumption and transfer shocks on initial conditions

variable	$IRF_{t+1}^{Y}\|v^{G}$			$IRF_{t+1}^{Y}\|v^{Z}$		
	estim.	std.	t-val.	estim.	std.	t-val.
$\tilde{\pi}_{t-1}$	-0.0024	0.0002	-10.20	-0.0076	0.0006	-12.87
$\tilde{Y}_{t-1}$	-0.0009	6.89E-06	-132.93	0.0011	1.72E-05	66.18
$\tilde{\iota}_{t-1}$	-0.0025	0.0002	-12.73	-0.0066	0.0005	-13.71
$\tilde{B}_{t-1}$	-6.57E-05	6.21E-07	-105.88	-0.0001	1.57E-06	-93.44
$\tilde{\tau}_{t-1}^{l}$	0.0002	4.43E-06	36.64	0.0003	1.11E-05	24.25
$\tilde{Z}_{t-1}$	-6.81E-05	3.54E-06	-19.21	-0.0002	8.89E-06	-25.82
$\tilde{G}_{t-1}$	0.0013	2.91E-06	451.00	5.14E-05	7.27E-06	7.08
$\tilde{a}_{t-1}$	-0.0001	2.14E-06	-59.49	-0.0004	5.38E-06	-75.55
$\tilde{\tau}_{t-1}^{c}$	2.73E-05	4.00E-06	6.83	4.38E-05	9.89E-06	4.42
π_{t-1}^{*}	0.0005	0.0005	1.08	0.0037	0.0012	3.17
Const.	0.1278	0.0001	1187.34	-0.0023	0.0003	-8.44
R^2	0.9305			0.6973		
$RMSE_{lin}$	0.0264			0.0660		
$RMSE_{mean}$	0.1000			0.1200		
obs.	60000			60000		

[16] While the structural, fiscal shocks in x_{t-1}^{L} may be relevant, they are also exogenous by design. Excluding exogenous variables should not affect other coefficients in this case. The structural shocks may be useful as controls. Though, as sample size and, thus, precision is not a limiting factor in this analysis I opt to omit as the share in variation explained by the structural shocks is fairly low.

Notes: Regressions of IRF^{Y}_{t+1} on initial conditions for a positive, one standard deviation shock to government consumption and transfers, respectively. Initial conditions are phrased as percentage steady state deviations as per the model set-up. $RMSE_{lin}$ is the in-sample root mean square error of the full linear model and $RMSE_{mean}$ is the RMSE for a mean model.

Table 2.7: Regression of IRFs on impact of output to consumption and labour tax shocks on initial conditions

variable	$IRF^{Y}_{t+1} \vert v^{\tau^c}$			$IRF^{Y}_{t+1} \vert v^{\tau^l}$		
	estim.	std.	t-val.	estim.	std.	t-val.
$\tilde{\pi}_{t-1}$	-0.0012	7.47E-05	-16.69	-0.0036	0.0002	-15.87
$\tilde{Y}_{t-1}$	0.0001	2.15E-06	47.31	0.0003	6.52E-06	44.88
$\tilde{\imath}_{t-1}$	-0.0008	6.17E-05	-13.32	-0.0026	0.0002	-13.85
$\tilde{B}_{t-1}$	-2.35E-05	1.99E-07	-118.19	-7.78E-05	5.96E-07	-130.64
$\tilde{\tau}^{l}_{t-1}$	4.44E-05	1.41E-06	31.38	0.0003	4.26E-06	67.69
$\tilde{Z}_{t-1}$	-2.34E-05	1.12E-06	-20.94	-6.50E-05	3.40E-06	-19.10
$\tilde{G}_{t-1}$	3.65E-06	9.29E-07	3.93	2.29E-05	2.78E-06	8.25
$\tilde{a}_{t-1}$	-3.15E-05	6.62E-07	-47.51	-8.54E-05	2.05E-06	-41.69
$\tilde{\tau}^{c}_{t-1}$	5.53E-05	1.29E-06	42.73	1.48E-05	3.94E-06	3.77
π^{*}_{t-1}	0.0013	0.0001	8.91	0.0038	0.0004	8.56
Const.	0.0045	3.42E-05	132.20	0.0198	0.0001	190.99
R^2	0.7271			0.6838		
$RMSE_{lin}$	0.0084			0.0254		
$RMSE_{mean}$	0.0160			0.0452		
obs.	60000			60000		

Notes: Regressions of IRF^{Y}_{t+1} on initial conditions for a positive, one standard deviation shock to consumption and labour taxation, respectively. Initial conditions are phrased as percentage steady state deviations as per the model set-up. $RMSE_{lin}$ is the in-sample root mean square error of the full linear model and $RMSE_{mean}$ is the RMSE for a mean model.

The first question is, "How relevant are the initial conditions in determining the effects of fiscal policy?". To assess this, I utilize a root mean square error comparison for in-sample predictions for the full linear model, $RMSE_{lin}$, and a version that only includes a constant term, $RMSE_{mean}$, which is equivalent to the steady state. Across both horizons and all fiscal shocks considered, the error of the full linear model is substantially lower than its constant counterpart. At the minimum, the *RMSE* is reduced by 32% for impulse responses to labour taxation shocks at four quarters. At the maximum, the *RMSE* is reduced by 74% for government consumption shocks on impact. Based on this, in a non-linear model, the initial conditions can be highly useful in pinning down the effects of fiscal policy.

The second, more general question is, "How do the effects of fiscal policy vary with the initial conditions?". Across the board, the coefficients of the impulse responses on the initial inflation and interest rate are negative for fiscal stimuli. That means that if either variable, inflation or interest, is below the steady state, then policy interventions are more stimulative. In a sense, this mirrors results of the Zero Lower Bound theory as in Woodford (2011) and Christiano, Eichenbaum and Rebelo (2011) on the interest rate side. They show that in periods of zero interest rates, the effects of fiscal stimulus are heightened and can be substantially larger.

Government consumption impulse responses on output are decreasing in output, while tax cuts and transfers are increasing. These results are in agreement with the analysis in Sims and Wolff (2013) and Sims and Wolff (2018a) on fiscal multipliers. In addition, it seems to be the case that all impulse responses to fiscal instruments are decreasing in debt. That implies that fiscal stimulus becomes more productive during periods when the government has low levels of debt and in the position to absorb the budgetary effects of stimulus.

In terms of scale, impulse responses on impact to government consumption shocks depend on the initial condition less than the other fiscal instruments. To illustrate this, an initial interest rate that is 1% below the steady state increases the effects of government consumption on output at impact by around 2% ($\frac{0.0025}{0.1278} * 100 = 1.96$). For consumption and labour taxation cuts, the predicted increase lies at around 18% and 13%, respectively. Similarly, if output is 1% below the steady state, then government consumption shocks are about 0.7% more effective, while tax cuts are about 2.25% or 1.5% less effective for consumption and labour taxation shocks. Furthermore, if inflation is 1% below the steady state, then government consumption shocks are roughly 1.9% more effective on impact. Tax cuts become 27.6% (consumption) and 17.9% (labour) more effective. Overall, for government consumption expenditures, the initial conditions are less relevant than for tax variables. However, it is important to note that business cycle conditions are never observed in isolation, and thus, government consumption impulse responses may vary across the cycle in a relevant way. For example, during the last 15 years, it is not atypical to observe periods of low interest rates combined with low inflation and low output. Based on this, the realized variation may be substantially larger.[17]

To sum up, this section argues that the initial conditions are particularly useful in pinning down the variation of impulse responses to fiscal interventions and can explain a range of estimates found in the previous literature. Regressing output IRFs to fiscal shocks on initial conditions, I find that fiscal policy is more effective at stimulating output in low interest rate, inflation and debt environments. I also find the government consumption multiplier is larger in recessions, while tax cut multipliers are larger in booms.

2.5.4 Historic path of policy effectiveness

This section shows how the regression results presented above can be applied to the actual business cycle conditions estimated over the sample data from Q1 1984 to Q4 2021.

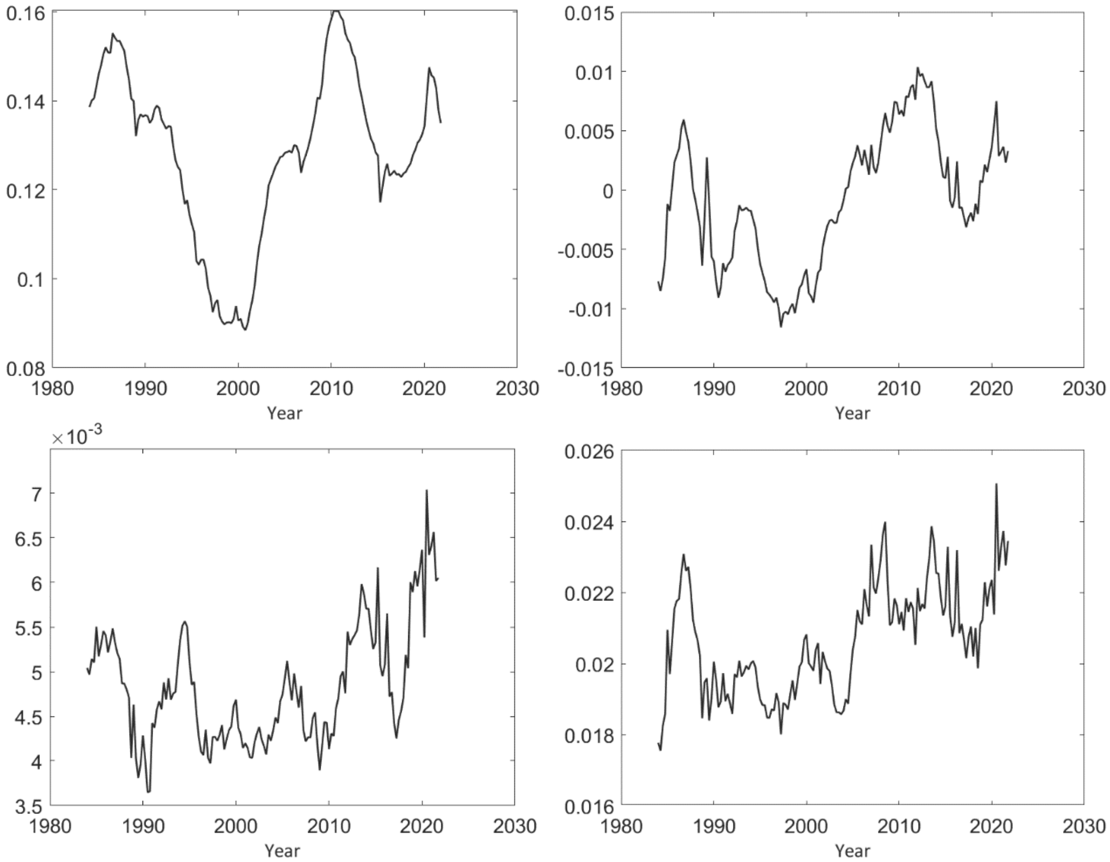
To do so, I construct the posterior mean estimates of the state vectors across the sample based on the conditional particle filter using parameter draws from the posterior. That includes estimates for all state variables of the DSGE, including inflation, output, interest rate and more. These state vectors are multiplied with the coefficients found in the regression exercise and finally averaged. The result is approximated paths for the effectiveness of the impulse responses for a given horizon and shock. Fig. 2.5 and Fig. 2.6 present the results.

Government consumption shows the clearest dynamic across the sample. In the mid to late 1980s, impulse responses of output to a one standard deviation government consumption shocks are most pronounced, climbing to above 0.15% at its peak on impact. After the start of the 1990s,

[17] A last note is on the relative importance of the initial conditions. In absolute terms, the coefficients of inflation and the interest rate are larger than for others variables like output and debt for all regressions. However, this does not mean that inflation and interest rates are more relevant in a typical business cycle situation. For example, the debt variable goes through very deep and protracted business cycles that may counteract smaller coefficients. To assess the realized impact of the initial conditions, one needs to consider both the scale of the coefficients and the spread of the variables.

government consumption stimulus becomes less and less effective, reaching its lowest values in and around 2000 at around 0.09%. After this period, the policy effectiveness almost doubles with the beginning of the financial crisis, reaching 0.16% in 2010. A similar increase, albeit lower in magnitude, is observed in 2020 during the pandemic.

Fig. 2.5: Paths of impact effect of fiscal policy around the cycle

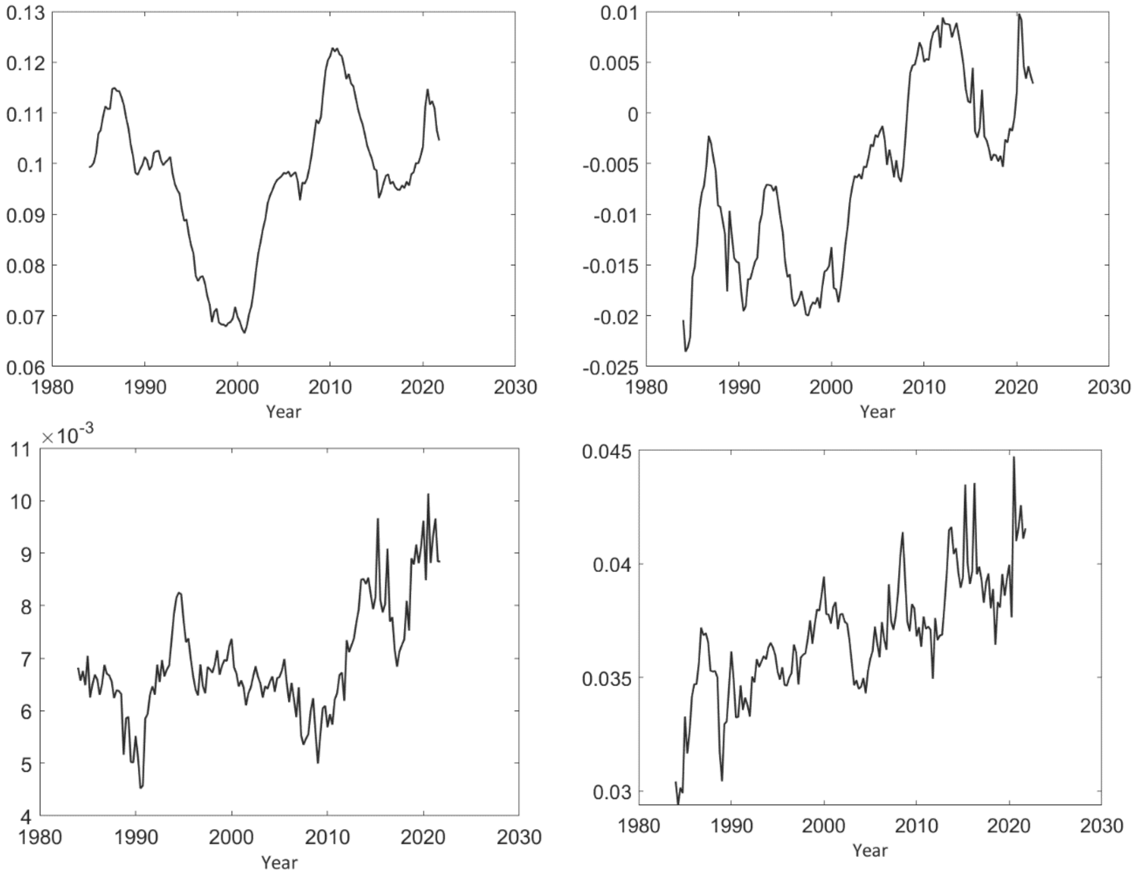


Notes: Constructed paths for IRF_{t+1}^{Y} in response to fiscal shocks for government consumption (upper left), transfers (upper right), consumption taxation (lower left) and labour taxation (lower right)

The tax variables go through less pronounced cycles overall. In both cases, the early 1980s are associated with slightly more effective impulse responses on impact, which are followed by a period of low effectiveness during the early 2000s. After the financial crisis, both tax variables show periods of increased effectiveness. Though, the timing is slightly different, and the persistence of this effect differs. For consumption taxation, effectiveness increases after 2000 and reaches a maximum in and around 2012 and decays afterwards temporarily. Instead, labour taxation effectiveness begins to increase before the start of the 2010s and remains higher more persistently. Both taxation variables spike in effectiveness during the Covid crisis. For the four-quarter impulse response, little changes for consumption taxation. For labour taxation, a trend to more effective policy arises over the whole sample.

Typically, transfers affect the economy by raising consumption and, thus, output. As argued in Leeper, Plante and Traum (2010), transfers, by themselves, are non-distortionary, and the effects of a transfer shock are mostly governed by how the fiscal shock is financed. For example, if it is tax financed, then a transfer shock may be followed by a reduction in government consumption and an increase in taxes. In this case, the effects of a transfer shock become less clear because it depends on the exact policy mix. Fig. 2.5 predicts that during the financial crisis and during the Covid crisis, transfers end up raising output at the mean. This stays the same even at longer horizons. However, before the year 2005, the effects of a transfer shock are estimated to be negative. Looking at section 2.5.5 , this coincides with both labour taxation and transfers becoming much more responsive to debt to curb the deficit. In essence, this is a policy mix more focused on financing shocks via taxes and less on raising debt.

Fig. 2.6: Paths of effect of fiscal policy at four quarters around the cycle



Notes: Constructed paths for IRF^{Y}_{t+4} in response to fiscal shocks for government consumption (upper left), transfers (upper right), consumption taxation (lower left) and labour taxation (lower right)

2.5.5 Policy gradients

The policy rules for the federal government and the central bank in this chapter include one novel feature: policy gradients may vary with business cycle conditions. This is the case as the rules are constructed as restricted, second-order Taylor approximation. Consequently, the gradients of the policy rules act as linear functions of the relevant business cycle conditions. In this section, I trace out the gradients of the fiscal rules with respect to output and debt across the sample. For the interest rate rule, I focus on constructing the time-varying estimates of the gradient of the interest rate to output growth and the percentage deviation of inflation to the inflation target.

In this application, I utilize the gradient of the Taylor rule described in the model section above. I pre-multiply the gradients of the policy rule with $(1-\rho_I)^{-1}$. The reason for this choice is that at the steady state, the two objects then have the familiar interpretation as being the two parameters ψ_y and ψ_π common to a lot of Taylor rules. Moving away from the steady state the second-order coefficients $\psi_{y,y}$, $\psi_{y,\pi}$ and $\psi_{\pi,\pi}$ start to bite:

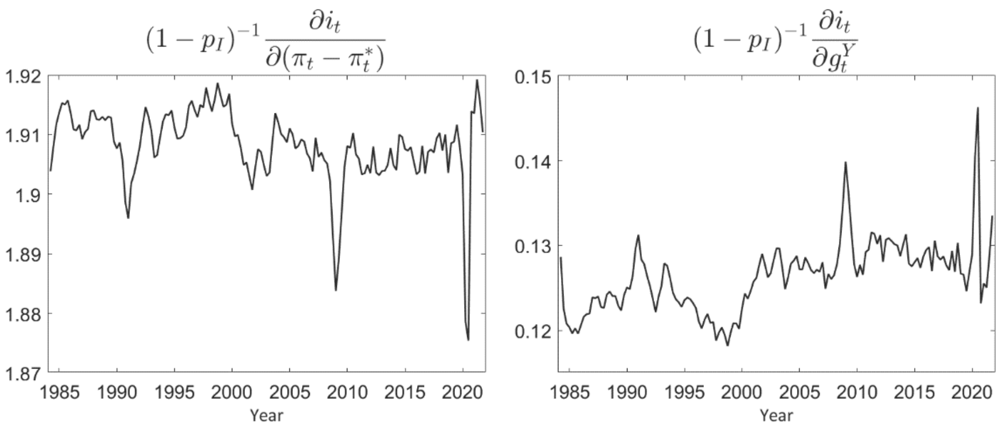
$$\psi_{y,t} = (1-\rho_I)^{-1}\frac{\partial i_t}{\partial(y_t - y_{t-1})} = \big(\psi_y + \psi_{y,y}(y_t - y_{t-1}) + \psi_{y,\pi}(\pi_t - \pi_t^*)\big),$$

$$\psi_{\pi,t} = (1-\rho_I)^{-1}\frac{\partial i_t}{\partial(\pi_t - \pi_t^*)} = \big(\psi_\pi + \psi_{y,\pi}(y_t - y_{t-1}) + \psi_{\pi,\pi}(\pi_t - \pi_t^*)\big),$$

where $\psi_{y,t}$ and $\psi_{\pi,t}$ are the pre-multiplied time-varying gradients. In a sense, the two objects, $\psi_{y,t}$ and $\psi_{\pi,t}$, can be interpreted as the expanded definition for ψ_y and ψ_π which allows them to change over the cycle. As $(1-\rho_I)^{-1}$ is constant across time and positively valued, any conclusion drawn from the adjusted gradients about correlation also applies to the actual gradients.

In order to implement this, I rely on the same sampling strategy as in the previous section. Based on posterior parameter draws, the mean state vectors are estimated across the sample. The state estimates are then multiplied with the corresponding elements of the posterior parameter vector to construct the gradient or objects of interest. Finally, the resulting estimates are averaged. The estimates for $\psi_{y,t}$ and $\psi_{\pi,t}$ are presented in Fig. 2.7.

Fig. 2.7: central bank policy rule gradients



Notes: Constructed paths for $\psi_{y,t}$ and $\psi_{\pi,t}$ in the central bank's Taylor rule across the sample. g_t^Y is the output growth rate otherwise also constructed as $(y_t - y_{t-1})$.

Firstly, both $\psi_{\pi,t}$ and $\psi_{y,t}$ show significant spikes around the time the US economy hits crisis. For example, at the beginning of the financial crisis, $\psi_{\pi,t}$ falls from up to 1.91 to around 1.88. At the same time, $\psi_{y,t}$ increases substantially from around 0.127 to above 0.14. While these are individually not substantial shifts, they do, however, suggest a shift in preferences by the central banks. Overall, the central bank became less concerned with ensuring inflation stays on target while becoming much more troubled about output growth. This is consistent with the observed policy measures during the crisis. The central bank released an unprecedented policy mix combining a low-interest rate strategy with quantitative easing. This policy mix was highly focused on controlling output, while inflation was of secondary concern. A similar pattern, and larger in magnitude, was observed during the Covid crisis. In general, this behaviour of increased responsiveness to output and decreased responsiveness to inflation is shared by all crises in the sample. To illustrate, one can detect local spikes during the early 1990s and early 2000s corresponding to the comparatively minor crises during those time periods.

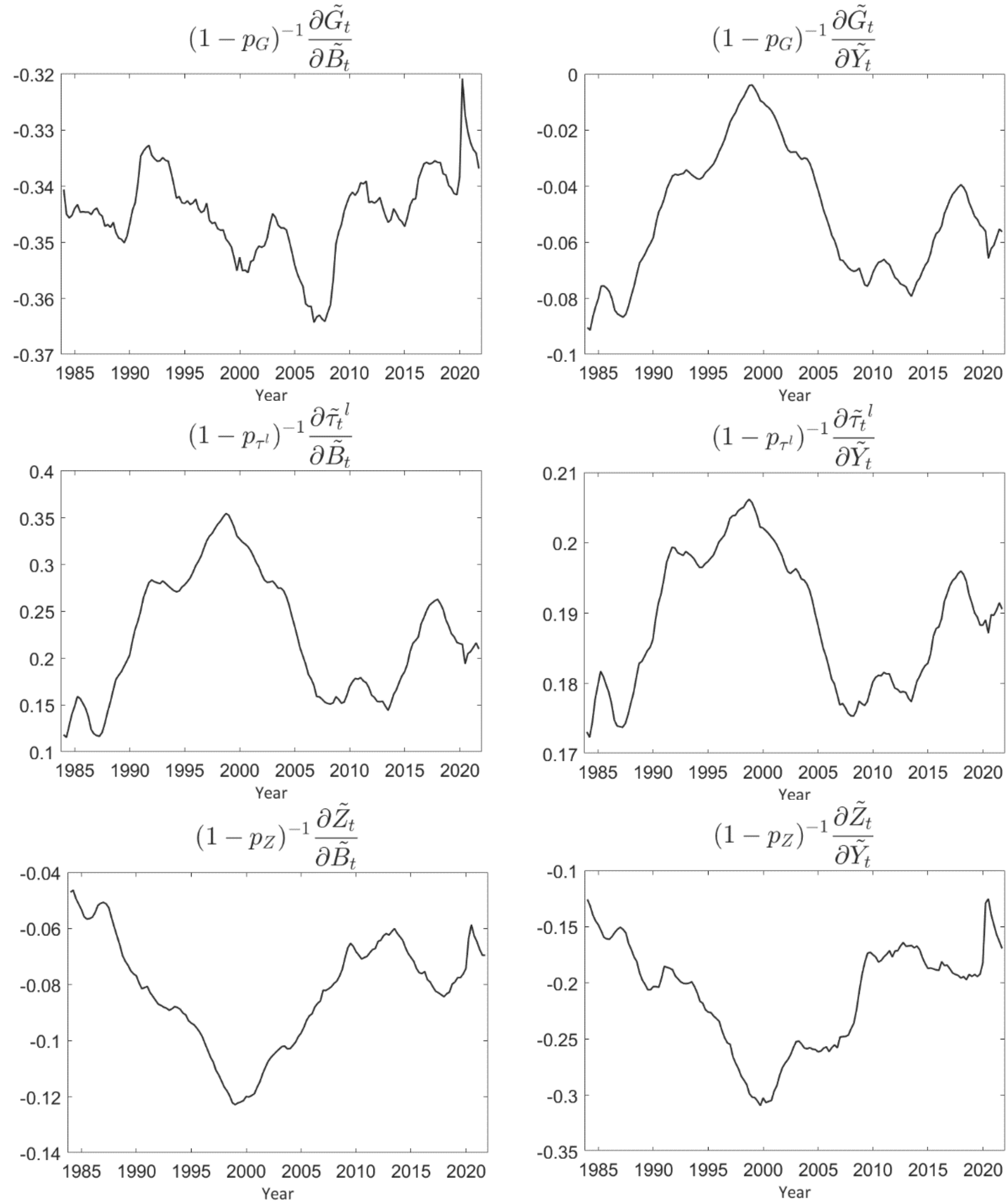
However, the persistence, scale and recovery of the gradient changes seem to differ from crisis to crisis. Mechanically, the reason for this is that the gradients are highly correlated with output growth. $\psi_{\pi,t}$ is positively correlated to output growth and $\psi_{y,t}$ negatively. Consequently, in boom phases, the central bank cares about controlling inflation and focuses less so on growth. As the economy moves away from a boom phase to a crisis, the central bank "switches" focus away from inflation to controlling output. Depending on the design of the economic crisis and how that translates to growth rates, responses in the gradients will be stark versus muted or persistent versus temporary.

In Fig. 2.7 the changes seemingly induced by crises seem to fade out comparatively quickly and introduce no long-term adjustments. The reason for this is that output growth, unlike output or output in terms of steady state deviations, features comparatively little persistence. So, while crises are easily recognizable in the data by large downward spikes, the spikes are usually temporary, followed by mildly negative or close to zero growth rates. Consequently, this explains why in this estimation, the changes in the gradient induced by economic crises are relatively short-lived.

A second result that can be inferred from Fig. 2.7 is that both $\psi_{\pi,t}$ and $\psi_{y,t}$ go through a mild mean adjustment similar to the US inflation rate. Overall, coming from the 80s and 90s, the 2000s and the financial crisis ushered in a period of persistently low inflation. This is reflected in the policy gradients via $\psi_{\pi,t}$ adjusting its mean downwards and $\psi_{y,t}$ adjusting upwards. Arguably, that seems to indicate that an overall shift in the policy rule took place during the shift in the interest rate mean.

Moving on to fiscal policy gradients, the gradients for the non-linear rules for government consumption, labour taxation and transfers are presented in Fig. 2.8. At the steady state, $\frac{\partial \tilde{G}_t}{\partial \widetilde{B}_t}$ is negative by design which implies that government consumption falls when debt increases. Here, $\frac{\partial \tilde{G}_t}{\partial \widetilde{B}_t}$ increases during periods of economic distress and, thus, government consumption becomes less responsive to debt. Large reductions in responsiveness can be seen in 2007 with the beginning of the financial crisis and, afterwards, with Covid as well. More muted reductions can be seen during the early 2000s crisis and the early 1990s recession as well. What this suggests is that in economic crises, the government's decision-making process for government consumption becomes substantially less concerned with controlling debt and, as such, paves the way for debt-financed expenditures.

Fig. 2.8: Government policy gradients around the cycle



Notes: Constructed paths for the rescaled gradients of the federal government rules across the sample.

For the remaining gradients, the key factor is that they are heavily correlated with the government debt, positively or negatively. For example, while $\frac{\partial \tilde{G}_t}{\partial \tilde{Y}_t}$ is negative across the sample, implying that government consumption increases in recessions as expected, the estimate gets close to zero in the 1990s when the debt level was very high. In practical terms, this implies that in economic downturns with high debt, the government consumption level will respond less to output than it would otherwise.

The labour taxation rate gradient to debt is also positively correlated with debt and changes quite substantially over time. Debt increasing in the 1990s coincides with the gradient of labour taxation to debt increasing in magnitude peaking in the late 1990s. In essence, as the debt level rises, labour taxation becomes more responsive to eventually force the budget back to the steady state. During the financial crisis, the labour taxation rate is in a period of relatively low responsiveness, making large stimulus packages possible. $\frac{\partial \tilde{\tau}_t^l}{\partial \tilde{Y}_t}$ is also positively correlated to debt. This suggests that the policy rule of the government in the 1990s implies much more stark increases in the labour taxation rate in response to above steady state output to balance the budget. In the early 2000s, the gradient begins to fall rapidly, and around the beginning of the financial crisis, the gradient is much smaller. $\frac{\partial \tilde{Z}_t}{\partial \tilde{B}_t}$ is negatively correlated to debt. Thus, transfers become more responsive to debt in the late 1990s to curb the deficit while being less responsive during the financial and Covid crisis. The same applies to $\frac{\partial \tilde{Z}_t}{\partial \tilde{Y}_t}$.

2.6 Conclusion

In this chapter, I propose a model that allows for the government and central bank to smoothly adjust their decision-making processes to the current state of the economy in a DSGE model. The model is estimated in its non-linear form using a fully Bayesian approach. Estimating DSGE models in their non-linear forms is a time-consuming effort even if vast computational resources are available. The research in this chapter combines pre-existing empirical advances to significantly cut down on the estimation time. The empirical framework itself is constructed based on key advances by Herbst and Schorfheide (2016), Jasra et al. (2010) and heavily borrows from the work of Buchholz, Chopis and Jacob (2021) on SMC samplers. Further, particular care was put into designing a code implementation that can keep up with the performance needs of the estimation by focusing on parallelization and vectorization wherever possible. Together, the estimation procedure reduces the estimation time from weeks to days by up to 94%, depending on the comparison basis. As a consequence, this chapter provides useful information on how to estimate non-linear models in a reasonable timeframe even on smaller machines.

Using the fully estimated model, it can be shown that the effects of fiscal policy vary significantly with the initial conditions of the economy and uncertainty is increased across the board if one does not condition on the steady state. To aid policymakers, I explore how the effects of fiscal policy relate to the initial conditions. To pin down this relationship, I prove that the impulse responses to a given shock are affine functions of the initial linear conditions. Based on this, a simple regression strategy can be used to quantitively express the relationship. The results show that all fiscal instruments are more stimulative in low interest rate periods and less effective in phases of above steady state debt. Overall, output impulse responses to tax cut shocks are estimated to be procyclical to output, consistent with Sims and Wolff (2018a), and government consumption is countercyclical.

I then combine the regression estimates with actual state estimates from historical US data from 1984 to 2021 to construct a time series for the time-varying effects of fiscal policy. Among all included fiscal instruments, government consumption goes through the most persistent cycles in its effectiveness. The results show that government consumption is estimated to have been most effective during the Covid and financial crises. Other instruments show less clear patterns but still show at least temporarily increased effectiveness during the zero lower bound period and Covid crisis.

The last contribution of this chapter comes from exploring what the non-linear government and central bank rules imply for their behaviour across the business cycle. I find that the interest rate rule is heavily determined by output growth. In periods of high output growth, the central bank is more concerned with controlling inflation and less concerned with adjusting to output growth. As the economy shifts into crisis, the central bank reduces its focus on inflation and shifts towards bringing output back onto target. For the fiscal rules, the key behaviour seems to be that gradients respond to the debt level. During the high debt period of the 1990s, labour taxation and transfers became increasingly responsive to debt and, therefore, adjust to ensure the financial stability of the federal government.

For future research, Gaussian process optimization seems promising. Gaussian process optimization is a Bayesian optimization technique typically applied to large-scale Machine Learning Systems and in non-linear model estimation. By design, Gaussian process optimization tends to be very performative in comparison to standard methods and variations of Bayesian optimization techniques for systems with latent states exist and are under development.

2.7 References

Amisano, G. and Tristani, O. (2007) "EURO AREA INFLATION PERSISTENCE IN AN ESTIMATED NONLINEAR DSGE MODEL", ecb.europa. Available at: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp754.pdf (Accessed: February 21, 2023).

Amisano, G. and Tristani, O. (2010) “Euro area inflation persistence in an estimated nonlinear DSGE model,” Journal of Economic Dynamics and Control, 34(10), pp. 1837–1858. Available at: https://doi.org/10.1016/j.jedc.2010.05.001.

An, S. and Schorfheide, F. (2007) “Bayesian Analysis of DSGE Models,” Econometric Reviews, 26(2-4), pp. 113–172. Available at: https://doi.org/10.1080/07474930701220071.

Andreasen, M.M., Fernández-Villaverde, J. and Rubio-Ramírez, J.F. (2017) “The pruned state-space system for non-linear DSGE models: Theory and empirical applications,” The Review of Economic Studies, 85(1), pp. 1–49. Available at: https://doi.org/10.1093/restud/rdx037.

Auerbach, A. and Gorodnichenko, Y. (2012) “Measuring the output responses to fiscal policy,” American Economic Journal: Economic Policy, 4(2), pp. 1–27. Available at: https://doi.org/10.1257/pol.4.2.1.

Baum, A. and Koester, G.B. (2011) The impact of fiscal policy on economic activity over the business cycle – evidence from a threshold VAR analysis, www.bundesbank.de. Deutsche Bundesbank. Available at: https://www.bundesbank.de/resource/blob/703584/e72e35658b3daf9e7286eb40f6962927/mL/2011-03-28-dkp-03-data.pdf (Accessed: March 17, 2023).

Blanchard, O. and Perotti, R. (2002) “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output,” The Quarterly Journal of Economics, 117(4), pp. 1329–1368. Available at: https://doi.org/10.1162/003355302320935043.

Boubaker, S., Nguyen, D.K. and Paltalidis, N. (2018) “Fiscal policy interventions at the zero lower bound,” Journal of Economic Dynamics and Control, 93, pp. 297–314. Available at: https://doi.org/10.1016/j.jedc.2018.01.048.

Buchholz, A., Chopin, N. and Jacob, P.E. (2021) “Adaptive tuning of Hamiltonian Monte Carlo within sequential Monte Carlo,” Bayesian Analysis, 16(3). Available at: https://doi.org/10.1214/20-ba1222.

Christiano, L., Eichenbaum, M. and Rebelo, S. (2011) “When Is the Government Spending Multiplier Large?,” Journal of Political Economy, 119(1), pp. 78–121. Available at: https://doi.org/10.1086/659312.

Creal, D. (2007) Sequential Monte Carlo samplers for Bayesian Dsge Models . Available at: https://people.bordeaux.inria.fr/pierre.delmoral/Creal_SMC_2007.pdf (Accessed: January 23, 2023).

Davig, T. and Leeper, E. (2009) “Monetary-fiscal policy interactions and fiscal stimulus.” Available at: https://doi.org/10.3386/w15133.

Del Moral, P., Doucet, A. and Jasra, A. (2006) “Sequential Monte Carlo samplers,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(3), pp. 411–436. Available at: https://doi.org/10.1111/j.1467-9868.2006.00553.x.

Demirel, U.D. (2021) “The short-term effects of tax changes: The role of State dependence,” Journal of Monetary Economics, 117, pp. 918–934. Available at: https://doi.org/10.1016/j.jmoneco.2020.07.002.

Drautzburg, T. and Uhlig, H. (2015) “Fiscal stimulus and distortionary taxation,” Review of Economic Dynamics, 18(4), pp. 894–920. Available at: https://doi.org/10.1016/j.red.2015.09.003.

Fazzari, S.M., Morley, J. and Panovska, I. (2014) “State-dependent effects of fiscal policy,” Studies in Nonlinear Dynamics & Econometrics, 19(3), pp. 285–315. Available at: https://doi.org/10.1515/snde-2014-0022.

Fernández-Villaverde, J. and Rubio-Ramírez, J.F. (2005) “Estimating dynamic equilibrium economies: linear versus nonlinear likelihood,” Journal of Applied Econometrics, 20(7), pp. 891–910. Available at: https://doi.org/10.1002/jae.814.

Fernández-Villaverde, J. et al. (2015) “Fiscal Volatility Shocks and Economic Activity,” American Economic Review, 105(11), pp. 3352–3384. Available at: https://doi.org/10.1257/aer.20121236.

Ferraresi, T., Roventini, A. and Fagiolo, G. (2014) “Fiscal Policies and Credit Regimes: A TVAR Approach,” Journal of Applied Econometrics, 30(7), pp. 1047–1072. Available at: https://doi.org/10.1002/jae.2420.

Fotiou, A. (2022) “Non-linearities in fiscal policy: The role of debt,” European Economic Review, 150, p. 104212. Available at: https://doi.org/10.1016/j.euroecorev.2022.104212.

Fotiou, A., Shen, W. and Yang, S.-C.S. (2020) “The fiscal state-dependent effects of capital income tax cuts,” Journal of Economic Dynamics and Control, 117, p. 103860. Available at: https://doi.org/10.1016/j.jedc.2020.103860.

Galí, J., López-Salido, J.D. and Vallés, J. (2007) “Understanding the effects of government spending on consumption,” Journal of the European Economic Association, 5(1), pp. 227–270. Available at: https://doi.org/10.1162/jeea.2007.5.1.227.

Gomes, S. et al. (2015) “Global policy at the zero lower bound in a large-scale DSGE model,” Journal of International Money and Finance, 50, pp. 134–153. Available at: https://doi.org/10.1016/j.jimonfin.2014.09.005.

Gomme, P. and Klein, P. (2011) “Second-order approximation of dynamic models without the use of tensors,” Journal of Economic Dynamics and Control, 35(4), pp. 604–615. Available at: https://doi.org/10.1016/j.jedc.2010.10.006.

Guerrieri, L. and Iacoviello, M. (2015) “Occbin: A toolkit for solving dynamic models with occasionally binding constraints easily,” Journal of Monetary Economics, 70, pp. 22–38. Available at: https://doi.org/10.1016/j.jmoneco.2014.08.005.

Herbst, E.P. and Schorfheide, F. (2016) Bayesian estimation of DSGE models. Princeton university press.

Huidrom, R. et al. (2020) “Why do fiscal multipliers depend on fiscal positions?,” Journal of Monetary Economics, 114, pp. 109–125. Available at: https://doi.org/10.1016/j.jmoneco.2019.03.004.

Jasra, A. et al. (2010) “Inference for lévy-driven stochastic volatility models via adaptive sequential Monte Carlo,” Scandinavian Journal of Statistics, 38(1), pp. 1–22. Available at: https://doi.org/10.1111/j.1467-9469.2010.00723.x.

Jerow, S. and Wolff, J. (2022) “Fiscal policy and uncertainty,” Journal of Economic Dynamics and Control, 145, p. 104559. Available at: https://doi.org/10.1016/j.jedc.2022.104559.

Jones, J.B. (2002) “Has fiscal policy helped stabilize the postwar U.S. economy?,” Journal of Monetary Economics, 49(4), pp. 709–746. Available at: https://doi.org/10.1016/s0304-3932(02)00113-7.

Kaplan, G. and Violante, G.L. (2014) “A model of the consumption response to fiscal stimulus payments,” Econometrica, 82(4), pp. 1199–1239. Available at: https://doi.org/10.3982/ecta10528.

Krajewski, P. and Szymańska, A. (2019) “The effectiveness of fiscal policy within business cycle – Ricardians vs. non-Ricardians approach,” Baltic Journal of Economics, 19(2), pp. 195–215. Available at: https://doi.org/10.1080/1406099x.2019.1609228.

Leeper, E.M., Plante, M. and Traum, N. (2010) “Dynamics of fiscal financing in the United States,” Journal of Econometrics, 156(2), pp. 304–321. Available at: https://doi.org/10.1016/j.jeconom.2009.11.001.

Mertens, K. and Ravn, M.O. (2014) “A reconciliation of SVAR and narrative estimates of tax multipliers,” Journal of Monetary Economics, 68, pp. S1–S19. Available at: https://doi.org/10.1016/j.jmoneco.2013.04.004.

Mountford, A. and Uhlig, H. (2009) “What are the effects of fiscal policy shocks?,” Journal of Applied Econometrics, 24(6), pp. 960–992. Available at: https://doi.org/10.1002/jae.1079.

Neal, R.M. (2012) MCMC using Hamiltonian Dynamics, arXiv.org. Available at: https://arxiv.org/abs/1206.1901 (Accessed: January 23, 2023).

Owyang, M. et al. (2013) “Are Government Spending Multipliers Greater During Periods of Slack? Evidence from 20th Century Historical Data,” American Economic Review, 103(3), pp. 129–134. Available at: https://doi.org/10.1257/aer.103.3.129.

Parker, J.A. (2011) “On Measuring the Effects of Fiscal Policy in Recessions,” Journal of Economic Literature, 49(3), pp. 703–718. Available at: https://doi.org/10.1257/jel.49.3.703.

Peren Arin, K., Koray, F. and Spagnolo, N. (2015) “Fiscal multipliers in good times and bad times,” Journal of Macroeconomics, 44, pp. 303–311. Available at: https://doi.org/10.1016/j.jmacro.2015.01.002.

Pfeiffer, J. (2018) A Guide to Specifying Observation Equations for the Estimation of DSGE Models. Available at: https://www.randomwalk.top/wp-content/uploads/2020/04/Pfeifer_2013_Observation_Equations.pdf (Accessed: January 23, 2023).

Ramey, V.A. and Zubairy, S. (2018) “Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data,” Journal of Political Economy, 126(2), pp. 850–901. Available at: https://doi.org/10.1086/696277.

Roeger, W. and in 't Veld, J. (2009) Ec.europa.eu. Available at: https://ec.europa.eu/economy_finance/publications/pages/publication13839_en.pdf (Accessed: 2022).

Romer, C.D. and Romer, D.H. (2010) “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks,” American Economic Review, 100(3), pp. 763–801. Available at: https://doi.org/10.1257/aer.100.3.763.

Schmitt-Grohé, S. and Uribe Martín (2004) “Solving dynamic general equilibrium models using a second-order approximation to the policy function,” Journal of Economic Dynamics and Control, 28(4), pp. 755–775. Available at: https://doi.org/10.1016/s0165-1889(03)00043-5.

Sims, E. and Wolff, J. (2013) “The output and welfare effects of government spending shocks over the business cycle.” Available at: https://doi.org/10.3386/w19749.

Sims, E. and Wolff, J. (2018b) “The output and welfare effects of government spending shocks over the business cycle,” International Economic Review, 59(3), pp. 1403–1435. Available at: https://doi.org/10.1111/iere.12308.

Sims, E. and Wolff, J. (2018a) “The state-dependent effects of tax shocks,” European Economic Review, 107, pp. 57–85. Available at: https://doi.org/10.1016/j.euroecorev.2018.05.002.

Smets, F. and Wouters, R. (2007) “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” American Economic Review, 97(3), pp. 586–606. Available at: https://doi.org/10.1257/aer.97.3.586.

Woodford, M. (2011) “Simple Analytics of the Government Expenditure Multiplier,” American Economic Journal: Macroeconomics, 3(1), pp. 1–35. Available at: https://doi.org/10.1257/mac.3.1.1.

Yang, Y. and Wang, L. (2015) An improved auxiliary particle filter for nonlinear dynamic equilibrium models, dynare.org. Available at: https://www.dynare.org/wp-repo/dynarewp047.pdf (Accessed: January 23, 2023).

Chapter 3

Forecasting with DSGE-VAR models using a model with rich fiscal rules

3.1 Introduction

In macroeconomics, there is a perceived split between models that are effective at forecasting and prediction purposes and models that deliver a structural interpretation of the economy, which allows for policy analysis. In particular, Pagan (2003) argues that there is a trade-off between the degree of theoretical coherence and the degree of empirical coherence, pitting purely theoretical models (e.g., hand calibrated frictionless DSGE models) against empirical reduced form models (e.g. VARs) and everything in between. With the turn of the century and the dawn of unprecedented computational opportunities available to researchers, the macroeconomic literature has quickly moved to bridge the gap by adapting the computationally and structurally complex DSGE models with features found in empirical models long known to aid forecasting performance. This chapter adds to this debate by evaluating the forecasting precision of the DSGE model presented in Chapter 2 and including a comparative analysis of more standard VAR and Bayesian VAR models for output and selected fiscal variables.

In this chapter, I attempt to provide additional evidence on the utility of DSGE models for forecasting using the DSGE-VAR framework. DSGE-VAR models, as developed by Del Negro and Schorfheide (2004), aim to resolve the split between VAR and DSGE models. This modelling framework is applied to the Chapter 2 model, which features a conventional model core heavily related to Amisano and Tristani (2010) and Smets and Wouters (2007). The key component of this model is a particularly detailed fiscal block that allows the government to react to the economy in a diverse way. In particular, it features two types of distortionary taxation – tax rates on labour income and consumption expenditures. Furthermore, it includes government consumption and transfers. Overall, the model is estimated in a data-rich environment. The forecasting accuracy for output, government consumption and federal debt obtained by this model is compared to standard VARs and Bayesian VARs with one and four lags. For the Bayesian VARs, the choice of prior fell on the popular Minnesota prior as developed by Litterman (1986) and Doan et al. (1984), which forms the ex-ante belief that macroeconomic series typically are well approximated by random-walk-like processes.

Previous results of the literature highlighted the importance of sample selection to forecasting accuracy of individual models. Therefore, this chapter estimates both rolling and expanding window versions over an expansive sample from 1954Q1 to 2021Q4 across a 1-step ahead and 4-step forecasting horizon. Further, forecasting performance is considered across the entire sample, and, in addition, two subsections from 1954Q1 to 1999Q4 and 2000Q1 to 2021Q4 are considered. This particular sample split reflects a significant shift in the dynamics after the beginning of the 2000s, owning to low-interest rates and high output volatility induced by the financial crisis and Covid crisis. The pre-2000s samples instead include a substantial share of the great moderation.

For the Chapter 2 model, the results show that the forecast accuracy for both output and fiscal variables improves almost uniformly when using a Bayesian prior to centre the reduced form parameter estimates. In general, neither the Minnesota prior nor the DSGE prior dominate the other across the board, in line with Gürkaynak, Kısacıkoğlu and Rossi (2014). The DSGE-VAR approach can provide substantial gains, especially at longer horizons, beating the Minnesota prior frequently for both output and fiscal variables. However, this may depend on the exact tuning of the DSGE-VAR approach, and mixed results do occur.

As has been shown in the literature, the smaller VARs with one lag outperform the larger VARs by substantial margins for output forecasting. One important advantage of the DSGE-VAR framework is that these larger VARs can become competitive if a DSGE prior is imposed tightly and can outperform significantly smaller VARs, which feature a fraction of the parameters. Consequently, for applied users, if there is value in studying larger models, then they can be made competitive by imposing tighter DSGE priors. This rule-of-thumb of smaller models being better continues to hold for government consumption but not necessarily for government debt. For forecasting government debt using a rolling window, the best fitting model is provided by

DSGE-VARs with fours lags, which improve upon its smaller counterparts by non-negligible margins.

A second contribution comes from a Zero Lower Bound (ZLB) estimation of the New Keynesian model. One of the critical advantages of DSGE models is that the economist can directly model complex dynamics and evaluate their performance. One such development is the inclusion of zero-lower bound constraints in DSGE models after the financial crisis. This chapter applies a standard zero lower bound constraint to its central Bank's Taylor rule. In this application, I use the solution strategy developed by Guerrieri and Iacoviello (2015) to solve the model and the filter proposed by Giovannini, Pfeiffer and Ratto (2021) is used to construct the likelihood. The ZLB model is estimated in an identical fashion using rolling and expanding window techniques. For output forecasting, the ZLB model performs reasonably well, outperforming the baseline VAR in many cases. However, in comparison to the DSGE-VAR approach, the performance seems to fall short. The ZLB is designed to incorporate an important dynamic of the financial crisis, but its forecasting performance is typically second to the DSGE-VAR framework. In comparison to output forecasting, the ZLB model proves quite useful for government consumption and debt forecasting, delivering the best-performing model more consistently.

Forecasting performance aside, the DSGE-VAR approach, as developed in Del Negro and Schorfheide (2004), offers additional insight about the structural nature of the data and allows for impulse response analysis, whereas the basic Bayesian VAR does not. The Bayesian prior used in the Minnesota approach is purely used to shrink the parameters to areas that are more likely ex-ante based on fundamental observations about macroeconomic data. It does not relate the reduced form shocks to the structural shocks of the model. In all typical circumstances, additional assumptions would be needed to recover the full structural model or to conduct impulse response analysis as described in Hamilton (1994). For the DSGE-VAR prior, that is not the case. Unlike the baseline VAR, the structural shocks of the DSGE model are identified and the model does not require additional assumptions to conduct impulse response analysis. As a by-product of the DSGE-VAR estimation, one receives an approximated posterior parameter distribution of the DSGE and, thus, has access to a fully identified structural model. For the Chapter 2 model, I trace out output impulse responses to government consumption shocks using the underlying

DSGE model for each sub-sample estimation conducted in the forecasting exercise.[18] The results show that impulse responses to government consumption shocks estimated on financial crisis data sets are more effective similar to the results in Chapter 2. Furthermore, government consumption is estimated to be more effective at increasing output during the mid-1980s and less effective in the late 1990s and early 2000s.

The structure of the chapter is as follows. Section 3.2 provides a general overview of the forecasting literature using DSGE models and the DSGE-VAR framework. Section 3.3 provides an overview of the Chapter 2 model, and Section 3.4 gives a detailed introduction to the empirical frameworks employed in this chapter, including the DSGE-VAR framework, Minnesota prior and the ZLB DSGE estimation strategy. Section 3.5 details the simulation strategy using the Metropolis Hasting algorithm, and Section 3.6 describes the data. The last sections present the results and the conclusion. The appendix features additional tables and some computational details.

3.2 Literature review

3.2.1 DSGE models

There is a large body of macroeconomic literature that focuses on exploring the forecasting abilities of DSGE models focusing on various aspects and comparing DSGEs to standard forecasting tools. Starting off with the general forecasting performance, in pioneering papers, Smets and Wouters (2003) and Smets and Wouters (2007) show that DSGE models can outperform the simpler VAR models in terms of forecasting performance. They achieve this by extending the Christiano, Eichenbaum and Evans (2005) model by including a number of economic frictions and exogenous driving processes aimed at improving the DSGE's fit to the

[18] Chapter 2 focuses on exploring a DSGE model where the effects of policy interventions are allowed to vary with the initial conditions of the business cycle. Here, I trace out the estimated effects of the linearized version of the same model. Therefore, the effects do not depend on the initial conditions but do depend on the sample selection of the windowed data set (i.e. the business cycle conditions over the sample).

data. Additionally, they also show that their model is competitive to Bayesian VARs using a Minnesota prior as in Litterman (1986) and Doan et al. (1984).

One key research question is how well these significant improvements generalize to other DSGE models. The initial forecasting success of Smets and Wouters (2003) and Smets and Wouters (2007) has spawned a considerable literature attempting to assess the forecasting performance of DSGEs. Results by Del Negro and Schorfheide (2013) show that an augmented Smets and Wouters (2007) model can perform markedly well compared to other professional forecasts like Blue Chip forecasts. Similarly, Berg (2016) finds that DSGE models can forecast comparatively well but may be poorly calibrated in some cases. However, other studies have shown that DSGE models may perform better or worse than the alternative models depending on the horizon or data series involved (see, e.g., Adolfson, Lindé and Villani (2007) and Christoffel, Coenen and Warne (2012)).

A more sobering view is provided by Edge and Gürkaynak (2011). They compare VAR, Bayesian VAR, Greenbook, Blue Chip forecasts and the Smets and Wouters (2007) model. They suggest limited advantages to using more complex models, such as DSGE and Bayesian VAR models, as they are often outperformed by simple models like constant or random walk models. Especially in terms of scale, the RMSE may be too large for the forecasts to be truly useful in policymaking. In a similar exercise, Gürkaynak, Kısacıkoğlu and Rossi (2014) find that among a large set of univariate and multivariate models, there does not exist a single best-performing forecasting method. In particular, smaller models may be more accurate at shorter horizons, and DSGE models favour longer horizons. Additionally, large-scale models are not preferable and often outperformed by small, low-dimensional AR and VAR models depending on the data set or variable. Equally, Chauve and Potter (2013) find simple linear or non-linear models can perform just as well as complex multivariate models, including DSGEs.

Wickens (2014) provides a survey of forecasting results using DSGE models by several institutions like the Reserve Bank of New Zealand, the IMF, Riksbank, the US federal reserve and other smaller forecasters. They show that DSGE models frequently forecast well, especially for longer horizons, but not better or worse than simpler time series models. They suggest that similarities in results are achieved by similar backwards-looking structures in both types of models. However, the forwards looking components of DSGE models, which depend on unknown exogenous processes, tend to be difficult to forecast.

The inclusion of several factors have been shown to improve DSGE model's performance, ranging from financial frictions (Cai et al. (2019), Gelfer (2019), Del Negro, Hasegawa and Schorfheide (2016)), a housing market (Kolasa and Rubaszek (2015)), an open economy (Gelfer (2021)), or stochastic volatility for the exogenous shock processes (Diebold et al. (2017)). In this chapter, I explore the inclusion of a rich fiscal structure into a DSGE model and analyse its importance for forecasting output and fiscal data.

3.2.2 DSGE-VAR models

An alternative approach to using the pure DSGE model for forecasting was developed by Del Negro and Schorfheide (2004) in the DSGE-VAR framework. Del Negro and Schorfheide (2004) achieve an efficient combination of the forecasting performance of VAR models while also retaining the causal analysis advantages DSGE models bring to the table. The DSGE-VAR methodology assumes that a VAR model is the data-generating process. However, Del Negro and Schorfheide (2004) utilize artificial data generated by the DSGE model to shift the parameters of the reduced form VAR into areas of the parameter space that are plausible ex-ante based on the DSGE model. The resulting advantage is that it allows for the estimation of the VAR reduced form parameters using economic beliefs specified for the DSGE. At the same time, it completely removes distributional restrictions and may reduce potential identification issues common to DSGE models. By doing so, the framework may ease up on some of the less empirically plausible modelling restrictions while maintaining the general covariate structure. Del Negro and Schorfheide (2004) show that DSGE-VAR models can provide robust forecasts comparable or better to VARs and BVARs.

The forecasting ability of DSGE models is also being explored through the DSGE-VAR framework. Ghent (2009) compares several standard real business cycle models through the use of the DSGE-VAR framework. They show that while the models include different structural characteristics, they perform relatively similarly. Equally, DSGE-VARs compare favourably to

VARs. Pop (2017) shows for Romanian data that their DSGE-VAR model is competitive with VAR models, beating VAR forecasts for inflation, real rate, and nominal interest rate forecasts but not real GDP. Furthermore, Gupta and Steinbach (2013) show that a small open economy DSGE-VAR model designed for South Africa performs well or better than comparable methods for forecasting output, inflation and a nominal short-term interest rate. In a DSGE-VAR model with a rich fiscal policy ruleset, Babecký et al. (2018) show that the DSGE-VAR model is preferable over VAR variants using a data density comparison. In terms of its usage of rich fiscal policy rules, this chapter is closely related to Babecký et al. (2018).

3.3 Chapter 2 DSGE model

For the estimation in this chapter, I use the model developed in chapter 2 and estimate a linearized version to simplify the estimation procedures. For the sake of brevity, here I focus on the fiscal and monetary policy rules as they are affected by the switch to linearization. For the remaining components of the model, I refer to Chapter 2.

The fiscal government features a budget constraint to ensure solvency. The government receives tax income based on labour, τ_t^L, and consumption taxation, τ_t^C, on their respective tax bases. On the expenditure side, it sets transfers, Z_t, and government consumption, G_t. Lastly, it has to refinance the one-period bond it gives out every period, which requires taking up new debt, B_t, and repaying last periods bonds with interest, $\frac{I_{t-1}B_{t-1}}{\pi_t}$. The following equation describes the government constraint:

$$\frac{\tau_t^C}{1+\tau_t^C}C_t + \frac{\tau_t^L}{1+\tau_t^L}\frac{1}{P_t}\int_0^1 W_t(i)L_t(i)di + B_t = Z_t + G_t + \frac{I_{t-1}B_{t-1}}{\pi_t},$$

where $W_t(i)L_t(i)$ is the labour income from good i, P_t is the current price level and C_t is current consumption.

For the fiscal instruments τ_t^L, Z_t and G_t the linear version of the fiscal rules is as follows:

$$\tilde{z}_t = p_z\tilde{z}_{t-1} + (1-p_z)\Big(k\mu_{z,Y}\tilde{Y}_t + \mu_{z,\pi}\tilde{\pi}_t + k\mu_{z,B}\widetilde{B}_t + \mu_{z,A}\tilde{A}_t\Big) + \sigma_z v_t^z, \quad v_t^z{\sim}N(0{,}1),$$
$$k = 1\ if\ \tilde{z}_t = \tilde{\tau}_t^l\ and\ else\ k = -1,$$

where $\tilde{z}_t$ is representative of the fiscal instrument in log steady state deviation form as indicated by the tilde. The parameters $\mu_{z,Y}$ and $\mu_{z,B}$ are assumed to be larger than zero for all z. The inclusion of k is done to ensure that the variables act to stabilize the budget based on the prior specification. That means that expenditure variables always decrease in response to increases in output or debt, while labour taxation does the opposite. Unlike standard rules, the fiscal instruments here are also allowed to respond to inflation and productivity to allow for richer behaviour. The fiscal instruments are perturbed by a random normal shock v_t^z with standard deviation σ_z. Consumption taxation receives a comparatively simple rule with an AR(1) process based on arguments in Leeper, Plante and Traum (2010) on the makeup of this tax rate in log steady state deviation form:

$$\tilde{\tau}_t^c = p_{\tau^c}\tilde{\tau}_{t-1}^c + \sigma_{\tau^c} v_t^{\tau^c}, \qquad v_t^{\tau^c} \sim N(0,1).$$

The Central Bank faces a standard Taylor rule where interest rates respond more than one-to-one to changes in inflation from its target. In addition, the CB also sets rates in response to output growth:

$$i_t = (1-\rho_I)\left(\bar{\pi} - ln(\beta) + \psi_y(y_t - y_{t-1}) + \psi_\pi(\pi_t - \pi_t^*)\right) + \rho_I i_{t-1} + v_t^i, \qquad v_t^i \sim N(0,\sigma_i^2).$$

Here, everything is phrased in log-terms. The log of the current interest rate, i_t, responds to the log of the lagged interest rate, output growth as constructed by the difference between todays and lagged log-output, $(y_t - y_{t-1})$, and the difference between the log of the inflation rate and the time-varying inflation target, $(\pi_t - \pi_t^*)$. To ensure stable inflation dynamics, ψ_π is larger than one and ψ_y is assumed to be larger than zero. v_t^i is the monetary policy shock.

For the ZLB variant, the model distinguishes between the shadow rate, i_t^{Shadow}, and the notional rate, i_t^N. The shadow rate is still always governed by the interest rate rule above. Therefore, it corresponds to the rule the Central Bank would set if it was not constrained by the ZLB. The notional rate, i_t^N, is the one that is observed in practice and holds the following law of motion:

$$i_t^N = max\{i^{min}, i_t^{Shadow}\},$$

where i^{min} is a small positive rate. Post-2008, the interest rate never dipped below zero. If one was to set $i^{min} = 0$, then the ZLB would never be reached in practice. In this implementation, I follow Boehl and Strobel (2022) by setting i^{min} such that the ZLB holds from 2009Q1 to 2015Q4. That means $i^{min} = 1.00053$, corresponding to a 0.053% per cent rate.

Furthermore, the CB has a non-constant inflation target governed by the following AR(1) process in log-terms:

$$\pi_t^* = (1 - \rho_\pi)\bar{\pi} + \rho_\pi \pi_{t-1}^* + v_t^\pi, \qquad v_t^\pi \sim N(0, \sigma_\pi^2).$$

What that means is that in particular scenarios, the CB may decide that the long-run target for inflation is not suitable and can deviate from this. Further, this means the CB changes its interest rate-setting behaviour based on the policy rule above. For example, if the inflation rate is significantly above target, this implies a rather strong interest response. However, if the target rises to accommodate, the Central Bank may opt for a much weaker response.

3.4 Empirical models

3.4.1 DSGE VAR

At its core, Del Negro and Schorfheide (2004) makes use of the idea that the $VAR(p)$ model may be used as an approximation to the moving average representation of the DSGE model. Unlike the $VAR(p)$ model, DSGE models impose relatively tight economic beliefs on the parameter distributions of the empirical model. Based on this, the first-order approximation of the DSGE model can help in defining plausible ranges for the $VAR(p)$ parameters, or in other words, it can be used as a prior. The key advantage of the methodology is that it allows the econometrician to learn both about the $VAR(p)$ parameters but also about the structural parameters that define the DSGE model.

To explore the estimation design of the DSGE-VAR framework, I begin by describing the canonical VAR framework. The set up for the standard $VAR(p)$ framework, which forms the basis for the approach in Del Negro and Schorfheide (2004), is as follows:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t, \quad u_t \sim N(0, \Sigma),$$

where y_t is a $(n \times 1)$ vector of observables that depend linearly on its past's values. The matrices Φ_i are of dimension $(n \times n)$ and contain the model parameters. u_t corresponds to the $(n \times 1)$, mean zero and normally distributed, reduced form error process with full covariance matrix Σ $(n \times n)$. This system can be rewritten in a more convenient way by utilizing an OLS-like representation. The system is characterised by two matrices Y and X that carry the current observations and the first p lagged observation vectors, respectively. Let Y be a $(T \times n)$ matrix where each row t corresponds to ${y_t}'$. Further, X is a $(T \times k)$ matrix where row t of matrix X is equal to the row vector $x_t' = [1, y_{t-1}', \dots, y_{t-p}']$. In this, k is equal to the row dimension of y_t, n, times the number of lags, p, plus one for a constant term $(k = 1 + np)$. Using this alternative representation, the system can then be written as:

$$Y = X\Phi + U,$$

where $\Phi = [\Phi_0, \Phi_1, \dots, \Phi_p]'$ and U includes stacked rows of u_t'. Using the matrix definition, one can construct the likelihood function:

$$p(Y|\Phi,\Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\ (-\frac{1}{2} tr\Big(\Sigma^{-1}(Y'Y - \Phi'\mathrm{X}'\mathrm{Y} - \mathrm{Y}'\mathrm{X}\Phi + \Phi'\mathrm{X}'\mathrm{X}\Phi)\Big),$$

where tr is the trace of a matrix, and U follows a joint normal distribution with covariance matrix Σ. Del Negro and Schorfheide (2004) proposes the following hierarchical structure for the joint prior of the $VAR(p)$ parameters and the structural DSGE parameters:

$$p(\Phi, \Sigma, \theta) = p(\Phi, \Sigma|\theta)p(\theta),$$

where θ is the vector of structural parameters that define the DSGE model together with its prior $p(\theta)$. The density $p(\Phi,\Sigma|\theta)$ describes a relationship between the reduced form $VAR(p)$ parameters and the structural DSGE parameters. The key idea is that based on the structural parameter, θ, the DSGE model defines moments that can be used to construct a prior for the VAR parameters. Here, let $\Gamma_{yy}^*(\theta)$, $\Gamma_{xy}^*(\theta)$ and $\Gamma_{xx}^*(\theta)$ denote the population moments of the DSGE model with parameter vector θ. The joined density is as follows:

$$p(\Phi,\Sigma|\theta) = c^{-1}(\theta)|\Sigma|^{-\frac{\lambda T+n+1}{2}} \exp\ \ (-\frac{1}{2} tr\left(\lambda T \Sigma^{-1}\left((\theta) - \Phi'\Gamma_{xy}^*(\theta) - \Gamma_{yx}^*(\theta)\Phi + \Phi'^{\Gamma_{xx}^*}(\theta)\Phi\right)\right),$$

where tr is the trace of a matrix. Conceptually, this likelihood can be compared to a likelihood of the $VAR(p)$ parameters Φ and Σ on an artificial data set simulated from the DSGE model.

Instead of using artificial data, which introduces unnecessary simulation noise, Del Negro and Schorfheide (2004) rely on the population moments. To complete the prior, Del Negro and Schorfheide (2004) introduce two things. Firstly, the parameter combination of λT in which λ governs the tightness of the DSGE prior by adjusting the number of artificial observations. For smaller values of λ and thus small artificial sample sizes, the posterior estimates are closer to the standard OLS estimates. Increasing λ, implies that the DSGE model gains influence over the $VAR(p)$ parameters as the size of the artificial sample increases. In the limit, as $\lambda \to \infty$ the DSGE-VAR approach reduces to estimating the DSGE model using a quasi-likelihood approach. Secondly, the shape of the prior distribution for Φ and Σ conditional on θ is normal and inverse Wishart, respectively:

$$\varSigma|\theta \sim IW(\lambda T \varSigma^*(\theta), \lambda T - k, n),$$
$$\varPhi|\theta \sim N(\varPhi^*(\theta), \varSigma \otimes \left(\lambda T \varGamma^*_{xx}(\theta)\right)^{-1},$$

where $\Sigma^*(\theta)$ and $\Phi^*(\theta)$ are defined as:

$$\varPhi^*(\theta) = \varGamma^*_{xx}(\theta)^{-1} \varGamma^*_{xy}(\theta),$$
$$\varSigma^*(\theta) = \varGamma^*_{yy}(\theta) - \varGamma^*_{yx}(\theta) \varGamma^*_{xx}(\theta)^{-1} \varGamma^*_{xy}(\theta).$$

The prior is proper under two conditions. Firstly, λT needs to be at least as large $k + n$ such that the inverse Wishart prior is defined. Secondly, the population moment matrix, $\Gamma^*_{xx}(\theta)$, needs to be invertible. If that is the case, then the two equations constitute a fully defined prior.

Based on the prior, the posterior distribution for the $VAR(p)$ parameters and the structural DSGE parameters is defined as follows:

$$p(\varPhi, \varSigma, \theta|Y) = p(\varPhi, \varSigma|\theta, Y) p(\theta|Y).$$

To simulate the posterior, both $p(\Phi, \Sigma|\theta, \mathrm{Y})$ and $p(\theta|\mathrm{Y})$ need to be characterized. Conditional on θ, sampling and evaluating the density $p(\Phi, \Sigma|\theta, \mathrm{Y})$ turns out to be straightforward as analytical results exist. Because of the choice of prior distribution for Φ and Σ, it can be shown that the posterior is also of the Inverse Wishart and Normal form:

$$\varSigma|Y, \theta \sim IW\left((\lambda + 1) T \tilde{\varSigma}(\theta), (1 + \lambda) T - k, n\right),$$
$$\varPhi|Y, \theta, \varSigma \sim N(\tilde{\varPhi}(\theta), \varSigma \otimes (\lambda T \varGamma^*_{xx}(\theta) + X'X)^{-1},$$

where $\widetilde{\Sigma}(\theta)$ and $\widetilde{\Phi}(\theta)$ are the MLE estimates of Φ and Σ on the augmented data set, which includes both artificial and actual observations.

To estimate the model, one has to simulate from the posterior of the structural parameters of the DSGE, θ. For $p(\theta|Y)$ the shape of the posterior is not known as is typical for DSGE models. Therefore, one has to rely on Markov-Chain-Monte-Caro (MCMC) techniques or similar simulation techniques. The MCMC algorithm is then applied to the distribution:

$$p(\theta|Y) \propto p(Y|\theta)p(\theta),$$

where $p(Y|\theta)$ can be constructed as follows:

$$p(Y|\theta) = p(Y|\ \Phi, \Sigma)p(\Phi, \Sigma|\ \theta)/p(\Phi, \Sigma|\ Y),$$

$$p(Y|\theta) \propto \frac{|\lambda T \Gamma^{*}_{xx}(\theta) + X'X|^{-\frac{n}{2}}|(\lambda+1)T\tilde{\Sigma}(\theta)\ |^{-\frac{(\lambda+1)T-k}{2}}}{|\lambda T \Gamma^{*}_{xx}(\theta)|^{-\frac{n}{2}}|\lambda T \Sigma^{*}(\theta)|^{-\frac{\lambda T-k}{2}}}.$$

In this case, I rely on the Metropolis-Hastings Algorithm utilized in Amisano and Tristani (2010). See section 3.5 for more detail. Using the generated draws from the distribution of $p(\theta|Y)$, one can then use the conditional posteriors for Σ and Φ to generate draws for the posterior $VAR(p)$ parameters. For λ I choose to set the hyperparameter equal to one and two. These reflect a similar range as tested in Del Negro and Schorfheide (2004) and perform well based on initial testing. Choosing larger values (>10) degrades forecasting performance in my testing. A more sophisticated approach to choosing the value of the hyperparameter is presented in Babecký et al. (2018). The main idea is to conduct a grid search to find the maximum of the log marginal likelihood and the associated λ. Their results show optimal λ's ranging from 1 to 3. Because the range of values includes the hand calibrated values chosen here and manual testing showed that the forecasting results are not majorly sensitive to fine-tuning the λ's in this range, I opted for manual calibration.

3.4.2 Minnesota prior

The following section gives a brief introduction to the Minnesota Prior for vector auto regressions developed in Litterman (1986) and Doan et al. (1984). For the implementation, I follow the more recent overview in Del Negro and Schorfheide (2010).

The motivation of the Minnesota prior can be based on the exposition in Todd (1984). When it comes to forecasting, personal expectations and prior beliefs can play a huge role in questions like model design, data selection and more. The standard VAR model in the frequentist format can be shown to reflect a diffuse set of prior beliefs on its parameters.[19] In practice, one may find that a diffuse or unassuming stance can often be at odds with the prior beliefs of forecasters and can be counterproductive for forecasting performance. The Minnesota prior aims to bridge that gap by incorporating reasonable prior information on macroeconomic data series. At its core, the Minnesota prior starts from the fundamental observation that typical macroeconomic data series show high persistence in their behaviour. In particular, simple random walk models or naïve forecasts almost always do well in forecasting comparisons despite or because of their simplicity. In the random walk model, the forecast of a specific observable, $y_{i,t+1}$, indexed by i, is just today's value:

$$y_{i,t+1} = y_{i,t} + \varepsilon_{i,t+1} \quad with\ E_t(y_{i,t+1}) = y_{i,t}.$$

Based on the observation that random walk models provide a good guess across a wide range of macroeconomic series, the Minnesota Prior incorporates this idea by shrinking the parameters of the VAR towards an independent random walk behaviour.

Within the Del Negro and Schorfheide (2010) framework, one adds dummy observations to induce the prior similarly to the DSGE priors. It can be shown that the inclusion of dummy observations leads to a conjugate prior of the normal, inverse Wieshart form. The dummy observations are added to the Y and X matrices below:

$$Y = X\Phi + U,$$

[19] The Ordinary least squares estimation of the VAR can be thought of as Bayesian VARs with a diffuse normal prior. From that perspective, the frequentist VAR is nested into the Bayesian VAR framework as a limit case.

where Y and U are $(T \times n)$ matrices, and X is a $(T \times k)$ matrix. Φ contains the VAR coefficient matrices as $\Phi = [\Phi_0, \Phi_1, \dots, \Phi_p]'$ and is of dimension $(k \times n)$. For more details on the exact set up of the linear regression system, see the previous section. To create the random walk behaviour, the parameters on the diagonal of Φ_1 receive a normal prior centred at 1, while off-diagonal elements are assumed to have a prior mean of zero. For the autoregressive matrices of higher order, the prior mean is set to zero. The standard deviations for those parameters decrease with the order of lags. This is included to avoid overfitting common to higher-order VARs.

The prior covariance matrix of the reduced form shocks is centred at a prior estimate of the covariance matrix. In addition, Del Negro and Schorfheide (2010) elaborate on two additional dummy observations for the intercept and to introduce correlation between model parameters. Firstly, the sums-of-coefficients dummy observations introduce the following mechanics: if $y_{i,t}$ is at its long-run mean or close to it, then the long-run mean is a good forecast for $y_{i,t+1}$. Secondly, the co-persistence dummy observations are introduced to reflect the prior that when y_t is at its long-run mean, then the long-run mean can serve as a good forecast for y_{t+1}.

To implement the Minnesota prior, Del Negro and Schorfheide (2010) rely on a set of hyperparameters: λ_1, λ_2, λ_3, λ_4, λ_5, $\bar{y}$ and $\bar{s}$. For this estimation, I set the hyperparameters as in Del Negro and Schorfheide (2010):

$$\lambda_1 = 0.1, \lambda_2 = 4, \lambda_3 = 1, \lambda_4 = 1 \textit{ and } \lambda_5 = 1.$$

$\bar{y}$ and $\bar{s}$ are calibrated to long-run estimates over the estimation sample.

3.4.3 ZLB DSGE

For the ZLB DSGE model, the estimation is conducted using the full likelihood as opposed to the quasi-likelihood approach utilized in the DSGE-VAR procedure. The fundamental reason for this is based on the way information about the DSGE is utilized in the DSGE-VAR approach. The DSGE-VAR approach summarizes the empirical properties of the DSGE by constructing first and second-order moments. For the VAR, this approach will capture all relevant moment information in a compact format. However, for the ZLB model, it is not immediately clear that the mean and covariance matrix will summarize the dynamics well. This is the case for two reasons. Firstly, it is not clear that the kink dynamics could be captured in the covariance matrix. Secondly, ZLB events are fundamentally rare. If one was to construct an unconditional covariance matrix, it would include little information about the different mechanics observed at the zero lower bound. An alternative strategy to utilise the DSGE-VAR approach could have been to replace the exact moments with sampled moments. This approach still suffers from the fact that ZLB events are rare and would introduce additional simulation noise. Therefore, I utilize exact likelihood methods in this chapter. To construct the likelihood, I use the filter developed in Giovannini, Pfeiffer and Ratto (2021). To solve the model, I utilize a variant of the algorithm for solving models with occasionally binding constraints (occbin) developed by Guerrieri and Iacoviello (2015). For more detail on the solution strategy, see the appendix. The posterior is then estimated based on the aforementioned filter and solution strategy using the Random Walk Metropolis-Hastings algorithm.

3.5 Bayesian Posterior simulation

To simulate the posterior of the structural parameters of the DSGE model in θ, I rely on a standard Random Walk Metropolis-Hastings Algorithm (RWMH), which works on a set of transformed parameters as in Amisano and Tristani (2010). The RWMH algorithm is a sequential sampler relying on the principles of Markov Chains. The sampler is initialized here based on a draw from an approximated posterior. Based on the initialization, the RWMH algorithm constructs a series of draws based on a proposal distribution. Each draw is either accepted or rejected based on an acceptance step. The chain then continues at the new parameter vector. Depending on the complexity of the posterior, the algorithm then converges to the posterior eventually.

The setup of the algorithm is as follows:

1. Generate proposal θ^* based on $q(\theta^*|\theta_{t-1})$, where $q(\theta^*|\theta_{t-1})$ is a multivariate normal distribution with a diagonal covariance matrix
2. Construct acceptance ratio: $a = \min\left\{\frac{k(\theta^*|Y)}{k(\theta_{t-1}|Y)}, 1\right\}$
    a. The function k corresponds to the kernel of the posterior:

b. $p(\theta|Y) \propto k(\theta|Y) = p(Y|\theta)p(\theta)$

3. Accept draw θ^* with $p = a$ and set $\theta_t = \theta^*$ or else reject and set $\theta_t = \theta_{t-1}$

In the case of DSGE models, a lot of elements in θ may be constrained to certain intervals. For example, habit persistence is typically constrained to the interval from zero to one. In this case, the proposal distribution $q(\theta^*|\theta_{t-1})$, can and will generate draws outside of the bounds of the prior. These draws will always be rejected. One way of avoiding this problem is working on a transformed set of parameters, ϕ, as in Amisano and Tristani (2010). ϕ is constructed using θ but is not bounded. For gamma and inverse gamma distributions, the transformation is a *log* transformation. Beta-distributed parameters are transformed using an inverse-sigmoid transformation. Normally distributed parameters are not transformed. To ensure the validity of the simulation, the kernel needs to be adjusted using the determinant of the Jacobian of the transformation:

$$p(\phi|Y) \propto p(Y|\phi)p(\phi) = p\big(Y\big|\theta(\phi)\big)p\big(\theta(\phi)\big)\left|\frac{\partial\theta}{\partial\phi}\right|.$$

The DSGE-VAR models and the ZLB-variant are re-estimated on every sub-sample using 10.000 MH draws. Overall, that constitutes to around 225 estimations per model, depending on the number of lags used.

3.6 Data

The Chapter 2 model is estimated using a comparatively rich data set which includes seven data series and spans from Q1 1954 to Q4 2021. The model features a similar fiscal apparatus to the Leeper, Plante and Traum (2010) model and, thus, shares similar variables. It includes output, debt, government consumption, a consumption tax rate, a labour tax rate, an inflation rate and an interest rate. The inflation rate corresponds to the implicit price GDP deflator, and the interest rate is the 3-Month Treasury Bill Rate. Nominal variables are deflated using an index constructed based on the GDP deflator using Q1 1990 as the base date. All variables but the inflation and interest rates are detrended using the same linear procedures as in Leeper, Plante and Traum (2010). This is done to induce stationarity in the variables. The resulting interpretation of these variables is as real, log steady state deviations. Log Interest rates and Log inflation rates are assumed to be directly measured by the DSGE. For more detail, see Chapter 2. Table 3.1 presents some detail on the used data series, their source and interpretation.

Table 3.1: Data overview

Variable	Type	Value	Source	Interpretation
τ^l	labour tax rate	rate	constructed as in Jones (2002); sourced from FRED and BEA	percentage deviatior from steady state
τ^c	consumption tax rate	rate	constructed as in Jones (2002); sourced from FRED and BEA	percentage deviatior from steady state
G	government consumption	real	FRED	percentage deviatior from steady state
B	government debt	real	FRED	percentage deviatior from steady state
Y	GDP	real	FRED	percentage deviatior from steady state
π	inflation	rate	FRED	log value
i	interest rate	nominal	FRED	log value

Notes: The table provides detail on the data set about the type, valuation, source, and interpretation of the variables. The federal reserve data base of the St. Louis FED bank is abbreviated as FRED, and the Bureau of Economic Analysis is abbreviated as BEA.

For the estimations in this chapter for both data sets, the variables are detrended for each individual sub-sample. For the expanding window estimations, the initial sample is set to start in Q1 1954 and includes 40 observations. The same applies to the rolling window estimation. All models are fully re-estimated on every subsample.

On a more general note, data construction is a crucial process in constructing DSGE models. The assumptions that are imposed on the data in the construction process can determine results when it comes to forecasting, structural analysis and other important topics. See Chapter 2 for a more detailed discussion on the choice of detrending options.

3.7 Forecasting performance

This section describes the forecasting results for the chapter 2 model using the DSGE-VAR and the ZLB model approaches. The analysis is broken down in the following way: It focuses on looking at the root mean squared forecast errors (RMSE) to evaluate forecasting performance. Here, the RMSE analysis focuses on forecasting percentage output, federal debt and government consumption deviations from a trend. RMSEs are evaluated in two ways. The first option is using a rolling window technique with a window size of 40 observations. The initial sample starts in Q1 1954. The second option that is employed is an expanding window. For out-of-sample performance, this chapter looks at two specific prediction periods. If the model parameters are estimated using a sample that includes observations of up to observation t, then $t+1$ and $t+4$ are considered for realistic forecasting horizons. $t+1$ corresponds to a one-quarter ahead prediction, and, in turn, $t+4$ corresponds to a full year. While one could consider alternatives and especially longer forecasting horizons, it is expected that accuracy will decrease across the board for all methods. The DSGE-VAR and ZLB approaches are then compared to the full set of chosen comparison models that includes the standard VAR and an alternative Bayesian VAR using the Minnesota prior. Especially the comparison to the Minnesota prior is relevant because it includes parameter shrinkage in a similar fashion. All models are evaluated using two variations: one with one lag order and one with four lags.

3.7.1 Forecasting output

Comparing frequentist VARs and Bayesian VARs

Table 3.2. shows that Bayesian VARs outperform the corresponding frequentist VAR of the same order at times quite substantially. For example, the Bayesian VAR(1) with a Minnesota prior outperforms the VAR(1) by 1.4% in a 1-step-ahead comparison and by 2.2% in a 4-step-ahead comparison for the rolling window. Similarly, the Bayesian VAR(1)s with a DSGE prior with $\lambda = 1$ $(\lambda = 2)$ improve on the VAR(1) by 7.9% (7.0%) and 2.3% (5.4%) for one and four step predictions. The results for the expanding window estimation are similar, with some differences. For the VAR(1) with the Minnesota prior, the gains at 1-step-ahead and 4-step-ahead are slightly

more muted at 0.6% and 0.06%. Equally, in the expanding window for predictions at four quarters head, the gains become more substantial for the DSGE-priors reducing the RMSE by 20.3% and 22.4% for $\lambda = 1$ and $\lambda = 2$, respectively.

Looking at VAR(4)s, the exact same thing holds, but the effect is magnified. For example, for the rolling window estimation, the inclusion of Bayesian priors can reduce the RMSE by anywhere from 32% to 87% depending on the forecasting horizon and prior. For the expanding window, this effect is smaller, reducing the RMSE by anywhere from 10% to 29%.

Table 3.2: Forecasting breakdown for output

		expanding window		rolling window	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.012	0.00	0.013	0.00
BVAR(1) Min	$t+1$	0.012	-0.64	0.012	-1.43
BVAR(1) $\lambda = 1$	$t+1$	0.011	-5.35	0.012	-7.85
BVAR(1) $\lambda = 2$	$t+1$	0.011	-4.95	0.012	-7.01
VAR(1)	$t+4$	0.030	0.00	0.027	0.00
BVAR(1) Min	$t+4$	0.030	-0.06	0.027	-2.19
BVAR(1) $\lambda = 1$	$t+4$	0.024	-20.33	0.027	-2.29
BVAR(1) $\lambda = 2$	$t+4$	0.023	-22.42	0.026	-5.39
VAR(4)	$t+1$	0.014	19.68	0.028	125.63
BVAR(4) Min	$t+1$	0.013	7.90	0.016	24.02
BVAR(4) $\lambda = 1$	$t+1$	0.012	5.75	0.019	52.66
BVAR(4) $\lambda = 2$	$t+1$	0.012	4.70	0.013	4.20
VAR(4)	$t+4$	0.038	27.55	0.225	729.01
BVAR(4) Min	$t+4$	0.033	9.53	0.029	4.83
BVAR(4) $\lambda = 1$	$t+4$	0.028	-6.02	0.088	223.10
BVAR(4) $\lambda = 2$	$t+4$	0.027	-10.06	0.032	16.27
ZLB	$t+1$	0.012	3.70	0.012	-0.48
ZLB	$t+4$	0.025	-15.84	0.026	-2.85

Notes: Breakdown of the forecasting performance of output for the DSGE model for the expanding and rolling window estimations. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the expanding and rolling window presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Here, the VAR(4)s profit substantially more from the Bayesian priors than the VAR(1). As mentioned before, the forecasting performance of the VAR(4) is substantially worse than the corresponding VAR(1). For example, for the rolling window, the RMSE of the frequentist VAR(1) is 0.013 and 0.027 for one and four-step ahead predictions. By comparison, the VAR(4) delivers an RMSE of 0.028 and 0.225 across the same horizons. By this simple comparison, the simpler VAR(1) is strictly preferable. Introducing a Bayesian prior in the form of the DSGE prior can improve forecasting performance substantially. For example, in the rolling window estimation, including a DSGE prior with $\lambda = 2$ in the VAR(4) can reduce the RMSE to 0.013 and 0.032 for the one and four-step horizons. While this does not improve upon the standard VAR(1), it does produce a comparable model in terms of forecasting performance. What Bayesian priors do is set up prior believes for the parameters included in the model based on ex-ante plausible ranges for the parameters. Depending on the weight of the prior, the parameter estimates are shifted towards these prior believes. Here, it turns out that the prior believes of the DSGE model, and the Minnesota prior are useful in reducing the RMSE and provide credible advantages in high-dimensional models like the VAR(4). The conclusion here is that larger models can be estimated and become useful forecasting tools by introducing Bayesian priors.

Minnesota Prior vs DSGE prior

So far, results have shown that Bayesian priors are useful across the board, but the key question is if either variant is preferable. In general, it seems to be the case that DSGE-VAR priors perform really well at longer horizons, that is at $t + 4$. Only for the rolling window estimation of the VAR(4) does the Minnesota prior beat all DSGE priors. This seems to suggest that the prior belief of the DSGE aide in forecasting performance more than the Minnesota prior in this application. In particular, reduced form DSGEs postulate dynamics of a stable, mean reverting system under the idea that in the medium or long run, the economy will return to the steady state. The Minnesota prior pushes the VAR estimates towards a random-walk-like dynamic. Consequently, it seems that for longer horizons, priors with a mean reverting property add a more useful dynamic to VAR models. Looking at $t + 1$ predictions, the results are similar. Overall, the DSGE prior proves to be a useful alternative to the Minnesota prior.

As an intermediate conclusion, the DSGE prior offer credible advantages for improving forecasting results. Several components of the DSGE methodology aid this in particular. Firstly, the DSGE model can help estimate larger VAR models, which are notoriously plagued by parameter uncertainty. Secondly, even in small VARs, the DSGE prior offers generally useful improvements in forecasting accuracy at all considered horizons. Thirdly, particularly at longer horizons, the DSGE priors show their full advantage by making use of the mean reverting property of the underlying DSGE. All in all, the DSGE prior offers a useful addition to the forecasting toolbox, and this makes the complicated construction of DSGE models worth the effort.

Forecasting using the ZLB framework

The last main component of this section focuses on the results of the Zero Lower Bound variation of the chapter 2 model. In comparison to the VAR(1) without a prior, the ZLB model forecasts output well but not uniformly better in both the rolling window estimation and the expanding window estimation. For the 4-step ahead predictions, the ZLB model outperforms the VAR(1) by 2.9% and 15.8% in the rolling and expanding window estimations, respectively. For the 1-step ahead predictions, the results are mixed. Most importantly, the ZLB approach is frequently outperformed by the simple DSGE-VAR approach.

There may be several explanations for this behaviour. Firstly, the model considered is a small to medium-scale new-Keynesian model. It lacks some of the characteristics that drove the dip in the real interest rate. Including a risk premium and capital market may model the data more accurately and improve the forecasting performance. Secondly, it seems that the combination of the filtering technique with the occbin solution strategy is subject to very nuanced details, e.g., the exact design of the occbin stepping algorithm. Holden (2021) shows that solution strategies to models with occasionally binding constraints may have one, infinite or zero solutions for a given state and shock vector. Consequently, there may be multiple expectational paths that agents may consider. For the estimation, this may provide a problem in that for a given shock vector and initial condition no unique solution may exist. Further, this may apply to multiple combinations of shock vectors and initial conditions. In practice, in this application, it seems that the filter developed in Giovannini, Pfeiffer and Ratto (2021) seems to frequently, but not always, cycle across different Zero Lower Bound durations. Arguably, a different implementation may lead to differing estimation results. Specifically, in this application, comparing the ZLB model to the simpler VARs with DSGE priors shows that the simplicity of the DSGE-VAR approach seems to outperform the added complexity of the ZLB model.

3.7.2 Forecasting Debt and Government Consumption

For government consumption and debt forecasts, most of the forecasting results for output also hold true. However, some differences to output forecasting can be found. The results can be found in Table 3.3 and Table 3.4.

Table 3.3: Forecasting breakdown for debt

		expanding window		rolling window	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.021	0.00	0.034	0.00
BVAR(1) Min	$t+1$	0.021	0.11	0.034	-0.64
BVAR(1) $\lambda = 1$	$t+1$	0.021	1.31	0.033	-4.33
BVAR(1) $\lambda = 2$	$t+1$	0.021	1.35	0.033	-4.11
VAR(1)	$t+4$	0.071	0.00	0.073	0.00
BVAR(1) Min	$t+4$	0.070	-0.15	0.072	-0.85
BVAR(1) $\lambda = 1$	$t+4$	0.068	-3.08	0.064	-11.94
BVAR(1) $\lambda = 2$	$t+4$	0.068	-3.57	0.064	-12.59
VAR(4)	$t+1$	0.042	100.01	0.039	13.68
BVAR(4) Min	$t+1$	0.022	5.81	0.033	-2.02
BVAR(4) $\lambda = 1$	$t+1$	0.024	15.31	0.032	-5.36
BVAR(4) $\lambda = 2$	$t+1$	0.020	-4.28	0.032	-6.73
VAR(4)	$t+4$	0.326	362.50	0.081	11.04
BVAR(4) Min	$t+4$	0.078	10.12	0.069	-4.85
BVAR(4) $\lambda = 1$	$t+4$	0.083	18.23	0.063	-14.00
BVAR(4) $\lambda = 2$	$t+4$	0.071	0.96	0.061	-15.71
ZLB	$t+1$	0.020	-5.32	0.032	-6.74
ZLB	$t+4$	0.059	-16.21	0.062	-15.40

Notes: Breakdown of the forecasting performance of debt for the DSGE model for the expanding and rolling window estimations. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the expanding and rolling window presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Table 3.4: Forecasting breakdown for government consumption

		expanding window		rolling window	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.012	0.00	0.043	0.00
BVAR(1) Min	$t+1$	0.012	-0.64	0.042	-0.35
BVAR(1) $\lambda=1$	$t+1$	0.011	-5.35	0.037	-13.18
BVAR(1) $\lambda=2$	$t+1$	0.011	-4.95	0.037	-12.60
VAR(1)	$t+4$	0.030	0.00	0.099	0.00
BVAR(1) Min	$t+4$	0.030	-0.06	0.097	-2.12
BVAR(1) $\lambda=1$	$t+4$	0.024	-20.33	0.090	-8.48
BVAR(1) $\lambda=2$	$t+4$	0.023	-22.42	0.089	-9.50
VAR(4)	$t+1$	0.014	19.68	0.078	82.39
BVAR(4) Min	$t+1$	0.013	7.90	0.043	0.29
BVAR(4) $\lambda=1$	$t+1$	0.012	5.75	0.045	4.98
BVAR(4) $\lambda=2$	$t+1$	0.012	4.70	0.040	-6.28
VAR(4)	$t+4$	0.038	27.55	0.147	48.85
BVAR(4) Min	$t+4$	0.033	9.53	0.103	4.48
BVAR(4) $\lambda=1$	$t+4$	0.028	-6.02	0.099	0.45
BVAR(4) $\lambda=2$	$t+4$	0.027	-10.06	0.092	-6.87
ZLB	$t+1$	0.012	3.70	0.035	-18.04
ZLB	$t+4$	0.025	-15.84	0.080	-18.87

Notes: Breakdown of the forecasting performance of government consumption for the DSGE model for the expanding and rolling window estimations. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the expanding and rolling window presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

For the two fiscal variables, the ZLB model seems to work better than for output. Looking at debt forecasts at the plus one horizon, the ZLB model improves upon the VAR(1) by 5.3% and 6.7% in the expanding and rolling window estimations. Further, in the four-quarter-ahead comparison, it comes out on top by reducing the RMSE by up to 16.2% (15.4%) over the VAR(1) for expanding (rolling) window. In the grand scheme of things, the ZLB variant and the DSGE-VAR approach still produce forecasts that are close together. For example, for the 1-step-ahead forecasts in the rolling window format, the DSGE-VAR(1) has an RMSE of 0.033 for $\lambda = 1$, while the ZLB comes in at 0.032, improving the forecast by 2.5%. Still, the ZLB model provides the best forecasts for government debt in all cases but the 4-step-ahead forecasts in the rolling window.

In general, it seems that for debt forecasts, more complex models can do better than smaller models, but this is not guaranteed. For example, for the rolling window forecast, the best fitting model at the 1-step-ahead forecast with one lag is the DSGE-VAR(1) with $\lambda = 1$ with an RMSE of 0.033. This is outperformed by the two VAR(4)s with the DSGE priors with both $\lambda = 1$ and $\lambda = 2$ and the ZLB model. Similarly, for 4-step-ahead forecasts, the best-performing model is the VAR(4).

Moving on from debt forecasting to government consumption, the ZLB model does not provide a uniform improvement over the VAR(1) for government consumption, but it does so in three out of four cases. In comparison to output forecasts, it also improves upon the VAR(1) more significantly. For example, at the 1-step-ahead horizon for the rolling window, it provides the best overall forecast improving on the VAR(1) by 18.04%. Further, for both windows at the 4-step horizon, the decreases in the RMSE are similarly substantial. However, for the expanding window, the overall best-fitting model class is still provided by the DSGE-VAR framework. While it seems that for debt, more complex models are advantageous for government consumption, this only holds for the ZLB framework and not for the larger VARS.

3.8 Forecasting performance on sub-samples

This section continues the type of analysis of the previous section but focuses on specific subsamples: The first sub-sample includes forecasting periods up to 1999 Q4 and therefore includes a large share of the Great Moderation. The second sub-sample includes the dot- com crisis, financial crisis and Covid crisis from 2000 Q1 to the end of the sample in 2021Q4. The forecasting tables for the rolling and expanding window estimations for output, debt and government consumption can be found in the appendix.

Previous results for the entire sample can be expanded unto the subsamples. In particular, the trend of smaller models being better forecasters continues to hold for output and government consumption. Additionally, Bayesian priors are generally useful in improving forecasting accuracy at short and long horizons. Though, there are cases depending on the specific sub-sample or window size, that standard VARs can outperform Bayesian VARs. Overall, the DSGE prior retains its edge over Minnesota prior in this estimation. But also, in this scenario, there are instances where the Minnesota prior achieves better forecasting accuracy. Further, Bayesian priors are still particularly useful in improving estimates of larger VARs and do so by quite substantial margins.

Generally, it seems that the initial periods of 1954 Q1 to 1999 Q4 generally produce better forecasts in terms of the RMSE for both expanding and rolling window estimations. This holds true for almost all model variants of the VAR. This is intuitive in the sense that the latter sample features ex-ante unlikely events being the financial crisis and the Covid crisis. Both crises introduce substantial forecasting uncertainty into the dynamic.

For output forecasting, one significant change is that for the post-2000s data set in the rolling window estimation, the DSGE-VAR approach with $\lambda = 1$ and lag order equal to one does not outperform the VAR(1) without a prior at four step ahead forecasts. However, increasing $\lambda = 2$ still generates prediction gains over the VAR(1) by 2.3%. In fact, this effect holds for all VAR estimations for both types of windows. Arguably, it seems to be the case that with short data sets that feature high uncertainty like the post-2000s period, tighter DSGE priors are more useful.

The ZLB model, which is designed to incorporate the post-2000 interest mechanics, does not improve output forecasting results consistently over the simpler VAR(1). For example, for the expanding window the ZLB model does improve over the VAR(1) at the 4-step horizon by 30% but fails to improve at all other horizons. For fiscal variables in the post-2000s, the ZLB approach can improve forecasting accuracy over the VAR(1), especially at longer horizons. For example, in comparison to the VAR(1), debt forecasts are improved by 1.46% and 1.5% at the 1-step-ahead and 4-step-ahead horizons in the rolling window estimations, while government consumption forecasts improve by 25.62% and 18.94% for the same horizons. But it does so inconsistently depending on the window type and data series, and the DSGE-VAR approach delivers easier and seemingly more robust results barring debt forecasts.

3.9 Stationarity Analysis

For modelling time series in general, one key component that the statistician has to analyse is if the data is stationary or non-stationary. A stationary series is governed by a dynamic system which does not depend on when it is observed. To illustrate, stationarity in its weakest form requires the unconditional mean and covariance matrix to be constant across time. A non-stationary series may violate either moment condition and, typically, could feature a trend, seasonality, or may be difference stationary.

In this chapter, all data sets feature non-stationary variables like output and government debt which trend upwards across time and arguably stationary variables like interest rates and inflation. However, for the actual estimation, all series that are non-stationary are detrended using a linear time trend. Consequently, the resulting sub-datasets feature series that show fairly constant empirical means across time. Ideally, when a $VAR(p)$ is fit to any of the sub-datasets, the resulting system ought to be stable. In practice, this chapter finds that this need not be the case for several reasons. Firstly, any of the estimation methods in this chapter do not put any hard constraints on the dynamics. Secondly, while any sub-sample may feature a stable mean across the entire sub-sample, locally, the data may be better explained by a non-stationary model. Thirdly, in macroeconomics, data sets are generally quite short as consistent modern data collection excludes pre-WW2 data typically, and the data collection is fairly low-frequency, focusing on quarterly and annual data. Rarely is it the case that rich monthly data over a long horizon is available. However, $VAR(p)$ models require a large number of coefficients to be estimated. To illustrate, $VAR(p)$ model features an observational vector, y_t, of dimension $(n \times 1)$ and p lags. The model includes n intercepts and $(p * n * n)$ autoregressive coefficients. The combination of generally small data sets and densely parameterized models can imply noisy estimates, which, together with the previous arguments, may explain non-stationary estimates.

In $VAR(p)$ models, non-stationarity can cause substantial issues when one looks at forecasting performance. To see that, in the following, I describe an alternative, canonical representation of the standard $VAR(p)$ in the form of the corresponding state space set up and construct predictions. In the state space setup, it is straightforward to address the topic of stationarity. The standard $VAR(p)$ framework is as follows:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + u_t, \quad u_t \sim N(0, \Sigma),$$

where y_t is a $(n \times 1)$ vector of observables that depend linearly on its pasts values. The matrices Φ_i for $i > 0$ are of dimension $(n \times n)$ and contain the model parameters. Φ_0 is a $(n \times 1)$ vector of constants. u_t corresponds to the $(n \times 1)$, mean zero, reduced form error process with full covariance matrix Σ $(n \times n)$. Instead of keeping track of several variables vectors, y_t, …, y_{t-p}, one can rewrite the above system using an expanded vector $\xi_t = [y_t' - \bar{y}', \dots, y_{t-p+1}' - \bar{y}']'$:

$$\xi_t = F\xi_{t-1} + v_t, \qquad v_t = [u_t, 0', \dots, 0'] \textit{ and } u_t \sim N(0, \Sigma),$$

$$y_t = \bar{y} + B\xi_t,$$

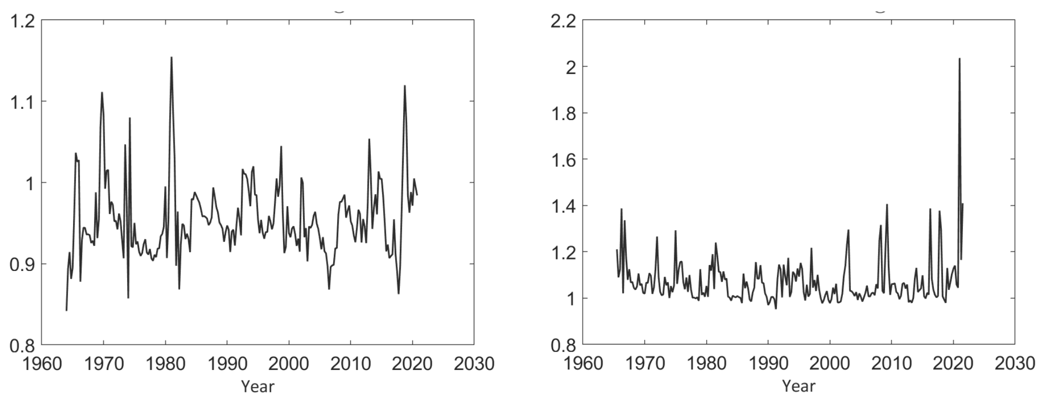
where $\bar{y}$ constitutes the unconditional mean of y_t if all eigenvalues of the matrix F are inside the unit circle. B is a $(n \times ((p-1) * n)$ matrix constructed as $B = [I_n, 0, .. ,0]$. The first n rows in F contain the autoregressive matrices Φ_1 to Φ_p side by side. After the first n rows, F includes a set of identity matrices to keep track of past values of $y_t - \bar{y}$. Conditional on ξ_{t-1}, the expectation of ξ_{t-1+h} and thus, the model forecast may be constructed as follows:

$$E(\xi_{t-1+h} | \xi_{t-1}) = F^h \xi_{t-1} \quad \textit{for } h \geq 1.$$

As a preliminary conclusion, the forecasts of the $VAR(p)$ are driven by the matrix F to the power of h. If the $VAR(p)$ is stable, then all eigenvalues of F are inside the unit circle and F^h will approach a zero matrix for larger forecasting horizons. As a by-product, it will generate stable predictions, which is consistent with the way the data sets are constructed in this chapter. However, if the largest eigenvalue is even slightly larger than or equal to one in absolute value, then the entire system becomes non-stationary and explosive. For larger forecasting horizons (i.e., $h = 4$), we may get diverging forecasts that are increasingly unrealistic. In practice, there is some nuance to this. If the largest eigenvalue of F is approximately one or only mildly larger than one, then the $VAR(p)$ can generate reasonable forecasts for appropriate choices of h.

Fig. 3.1 shows the largest absolute eigenvalue across time of the $VAR(1)$ and $VAR(4)$ estimations using a rolling window. The $VAR(4)$ estimation combines the previously mentioned estimation problems of a large, densely parametrized model, which features seven constants and 196 autoregressive parameters with the comparatively small samples size of the rolling window technique. Therefore, it's a prime candidate for potential instability. Across time the largest eigenvalue is consistently larger than one and frequently quite substantially so. For example, the inclusion of the Covid crisis causes the maximum absolute eigenvalue to spike substantially above 2. Consequently, as F^h is explosive, the forecasting accuracy degrades in comparison to the much smaller $VAR(1)$, which features eigenvalues much closer to one. Unstable or temporarily unstable models are usually a cause for concern of, for example, improper variable scaling. I test for this by estimating on a rescaled data set by using several rescaling techniques (standard deviation, mean absolute deviation). The results appear invariant to scaling.

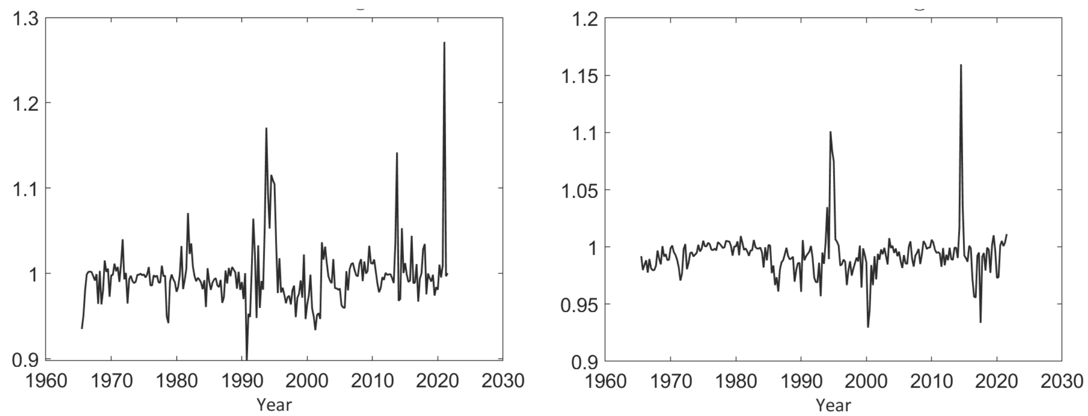
Fig. 3.1: Paths of maximum absolute eigenvalues for VARs



Notes: Maximum absolute eigenvalues figures for $VAR(1)$ (left) and $VAR(4)$ (right). The date indicates the end date of the sample. The eigenvalues are constructed based on the matrix F

Ideally, one would shrink the eigenvalues of the $VAR(4)$ towards one or below, which would be consistent with the data set construction. This is something that Bayesian priors excel at. In particular, the DSGE-VAR approach proposed by Del Negro and Schorfheide (2004) shrinks the parameters of the $VAR(p)$ to the maximum likelihood estimates of the DSGE using a quasi-likelihood based on a VAR of the same order. The parameter λ governs how dogmatically the DSGE dynamics are imposed on the VAR. The larger the value for λ is, the more strongly the DSGE dynamics are imposed. Crucially, DSGEs feature dynamics which are universally useful in this application. Standard reduced-form DSGE models are stable by design as all eigenvalues are assumed to be inside of the unit circle based on the solution technique. Therefore, any $VAR(p)$ is shrunken towards a model that is purely stable. The result is that the largest absolute eigenvalue decreases the stronger the DSGE is imposed. Fig. 3.2 below shows the maximum eigenvalues across time for the $DSGE-VAR(4)$ for λ equal to one and two. The first result is that for both $DSGE-VAR(4)$s the maximum eigenvalues are substantially smaller and closer to one than for the $VAR(4)$. Further, the maximum eigenvalues resemble the $VAR(1)$ much more closely. Lastly, as λ is increased from one to two and the DSGE prior is imposed more, the distribution of the maximum absolute eigenvalue becomes much tighter and more stable. Similar stability results can be obtained for the Minnesota prior as well. However, the parameters are shrunken towards the random walk hypothesis as opposed to the stable system set-up of DSGEs. This may or may not be a desirable property depending on the data set. Summing up, parameter shrinkage towards non-explosive models can be an important component of constructing forecasts. The $DSGE-VAR(p)$ methodology succeeds in creating more stable models based on fundamental economic observations, even in high dimensional models.

Fig. 3.2: Paths of maximum absolute eigenvalues for DSGE-VARs



Notes: Maximum absolute eigenvalues figures for $DSGE-VAR(4)$ with lambda equal to 1 (left) and $DSGE-VAR(4)$ with lambda equal to 2 (right). Date indicates the end date of the sample. The eigenvalues are constructed based on the matrix F

3.10 Impulse response analysis

As the structural shocks of DSGE models are fully identified by assumption, the DSGE-VAR approach allows the researcher to conduct impulse response analysis on the underlying model without any further steps. During the estimation procedure of the DSGE-VAR, one obtains posterior estimates of the VAR parameters but equally of the structural DSGE parameters. These posterior estimates can then be used to construct impulse responses.

As a by-product of the window estimations in the previous sections, posterior estimates of the structural parameters were constructed across a lot of subsamples. In this section, I utilize these posterior estimates and their distribution across time to explore if and how impulse responses change across time. The main emphasis of the analysis is on the impact of government consumption. The results below are obtained using data from Q1 1954 to Q4 2021.

Starting out, the linearized, structural DSGE model is typically governed by the following type of equation set based on the Klein (2000) set-up for the state system:

$$A(\theta)\begin{bmatrix} x_{t+1} \\ E_t y_{t+1} \end{bmatrix} = B(\theta)\begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \end{bmatrix}, \qquad v_t \sim N\big(0, \Sigma(\theta)\big),$$

where x_t is a $(k \times 1)$ vector of model states observed at time t. Here, x_t and y_t do not necessarily corresponds to the VAR variables and are defined as new vectors. Typically, the vector x_t has a row dimension that is larger than the number of observables that the model features. For example, the Chapter 2 model features 14 states but just seven observables. The vector y_t is a $(m \times 1)$ vector of endogenous, non-predetermined variables. The vector v_t is $(l \times 1)$ structural shock vector with covariance matrix $\Sigma(\theta)$. This structural shock vector, unlike the reduced form VAR shock vector, is assumed to be fully identified. The last components are the matrices $A(\theta)$ and $B(\theta)$ of size $((k+m) \times (k+m))$. The matrices are constructed by differentiating the full non-linear model at the steady state with respect to variables dated t and $t+1$. Notably, the matrices $A(\theta)$ and $B(\theta)$ are constructed using the structural parameter vector θ. Applying the Klein (2000) solution strategy allows for the construction of the reduced form system for the model variables. The methodology works via decomposing the full system into stable or backward-looking components like capital accumulation and unstable or forward-looking components like inflation. Roughly speaking, the unstable components receive their forward iterated solution, which is constrained to a stable solution. Based on this, the stable components receive their backwards solution. The solution is characterized by the following set of equations:

$$x_t = P(\theta)x_{t-1} + v_t, \qquad v_t \sim N\big(0, \Sigma(\theta)\big),$$

$$y_t = F(\theta)x_t.$$

For the construction of impulse responses, x_t is the main object of interest as it carries the dynamics across time. This vector depends linearly on itself in the previous period but also on a structural and identified shock vector v_t with covariance matrix $\Sigma(\theta)$. The relationship between the current and past state vectors is governed by $P(\theta)$. Unlike the VAR framework, the mapping between the structural parameter vector and the reduced form system is non-linear for DSGE models.

If a specific shock, v_t, occurs today, then the impact on future model states, x_{t+h}, is governed by the transition matrix $P(\theta)$. In particular, the expectation of x_{t+h} conditional on the shock, v_t, initial condition, x_{t-1}, and parameter vector, θ, at some horizon h is as follows:

$$E_t(x_{t+h}|x_{t-1}, v_t, \theta) = P(\theta)^{h+1} x_{t-1} + P(\theta)^h v_t.$$

To construct the impulse response, one compares the setting in which the shock occurred as above to a counterfactual where it did not. In the latter case, v_t is equal to zero in all entries. Formalizing it delivers the following equation:

$$IRF(h) = E_t(x_{t+h}|x_{t-1}, v_t, \theta) - E_t(x_{t+h}|x_{t-1}, v_t = 0, \theta) = P(\theta)^h v_t.$$

One crucial property of this system is that in linear modelling, the impulse responses are completely independent of the initial conditions, as can be seen by x_{t-1} dropping out. A by-product of this is that one can view the impulse responses relative to the steady state. At the steady state, the system, by definition, does not transition and $P(\theta)x_{steady\ state}$ is always zero. Consequently, one may interpret $IRF(h)$ as the steady state deviation caused by perturbing the model by v_t.

To construct posterior mean estimates of the impulse responses for DSGE models, the following procedure is followed. Firstly, draws are sampled from the posterior distribution of θ and the reduced form matrices $\Sigma(\theta)$ and $P(\theta)$ are constructed. Secondly, impulse responses are drawn, and thirdly, the impulse responses are averaged to approximate the mean posterior impulse responses. In this application, impulse responses are conducted based on 250 draws from the posterior of θ. The size of the government consumption shock is chosen to be equal to one standard deviation of the structural government consumption shock. As a result, the impulses below may be interpreted as the percentage steady state deviation of the variable of interest in response to a one standard deviation shock to government consumption. In this section, I prefer to use the underlying DSGE model estimated using the DSGE-VAR quasi-likelihood technique.

One advantage is that the underlying DSGE is directly comparable to the estimation in chapter 2 and can be seen as a linearized, quasi-likelihood version of it. Alternatively, one can also use the estimated DSGE matrices to identify the VAR, but this requires additional assumptions.

To explore how the effectiveness of government consumption stimulus changes across time for prediction purposes, I trace out mean posterior impulse responses of output across all individual sub-samples for the $DSGE - VAR(1)$ estimation with λ equal to 2 in the rolling window setting. The rolling window estimations are chosen because they will deliver a measure of the time variance of the posterior estimates as opposed to the expanding window. Fig. 3.3, Fig. 3.4, and Fig. 3.5 present slices of the mean impulse responses on impact, at four quarters, and at eight quarters. The x-Axis represents the corresponding end date for the estimation sample. Therefore, the first plotted observation has its start date in 1954 and ends in 1964. On impact, the results show that a one standard deviation increase in government consumption increases output by about 0.052 per cent on average relative to the steady state. This estimate is of a similar scale to Leeper, Plante and Traum (2010) and chapter 2, though lower. Furthermore, after four quarters, the impact has declined to 0.0112, and after an additional four quarters, the mean impact sits at 0.0048.[20]

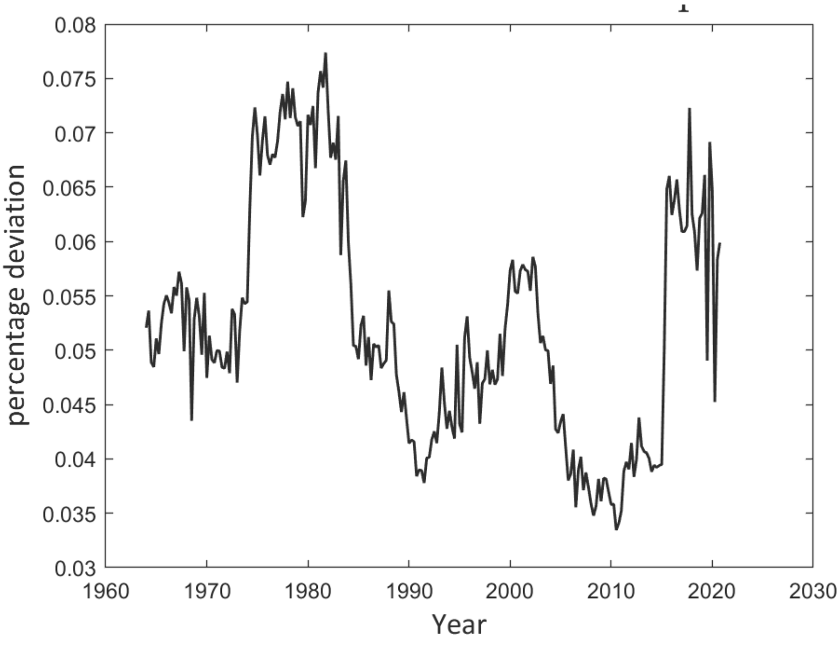
On impact, the mean of the impulse responses shows a significant time-varying component. To illustrate, the window estimation predicts that samples that have end dates in the mid-seventies to mid-eighties and include the previous ten years average significantly higher at 0.07. This is followed by a period of comparatively lower impulse responses until samples with end dates in the early 2000s. One particularly important period of modern economic history is the financial crisis of 2008. During the financial crisis, fiscal policy experienced somewhat of a renaissance. Monetary policy experienced constraints in the form of the Zero Lower Bound. Fiscal policy, however, is not limited by anything but the government's will to legislate and the government's

[20] The lower estimates very likely arise as a by-product of the change in model and estimation technique. In chapter 2, the second order approximation of the model is taken to the data, while here the linear DSGE is used to inform about a-priori likely VAR parameters. Additionally, unlike the full DSGE approach, the DSGE-VAR approach is a quasi-likelihood approach to estimating the DSGE model. How accurate the quasi-likelihood approximation is, is increasing in the number of lags and lambda. Here, the choices of the number of lags and λ reflect a very loosely imposed DSGE. The appendix presents estimation results for the model parameters averaged across the estimations and shows that some core economic parameters and policy parameters are estimated quite differently, and parameter uncertainty is substantially higher. Arguably, this justifies the difference in results.

debt obligations. As a by-product of the crisis, policymakers "rediscovered" fiscal policy tools and created enormous fiscal stimulus packages in most Western economies. One particular question of interest is how the effectiveness of government consumption expenditures changes in periods of low, if not nearly zero, interest rates. In this estimation, sub-samples that include a significant share of the financial crisis have end dates after 2015. These sub-samples show impulse responses of output to government consumption on impact that are far above average at around 0.065. This indicates that government consumption expenditures were indeed more productive in the financial crisis samples.[21]

For the impacts at four and eight quarters, the time-varying component is significantly lower and muted. This is to be expected as the impact of the spending shock gradually tapers out. Interestingly, the financial crisis data sets feature several mean responses that are fairly high but also feature a lot of uncertainty. Furthermore, at the longer horizons of four and eight quarters, the samples include the late 60s and early 70s, deliver the highest impact on average.

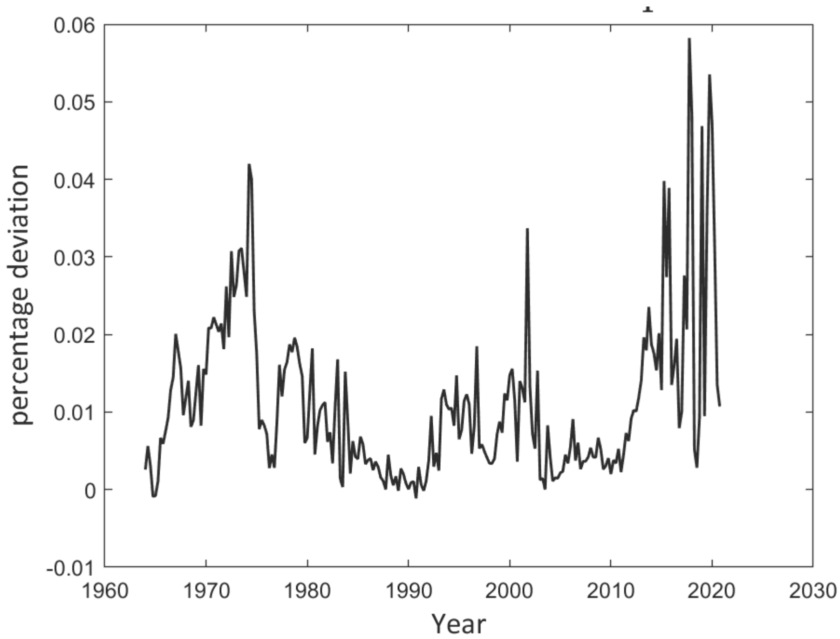
Fig. 3.3: IRF of Y to G shock on impact



[21] Similar to the results here, chapter 2 predicts that government consumption shocks are more productive on impact during the financial crisis. Furthermore, the results show that government consumption stimulus is in a period of high productivity from the mid-1980s to the mid-1990s. By comparison, this period of high effectiveness is estimated to be slightly earlier here with sample end-dates up to around 1985. However, both estimations predict low effectiveness during the late 1990s and 2000s. On a last note, chapter 2 predicts overall larger output responses than are estimated here.

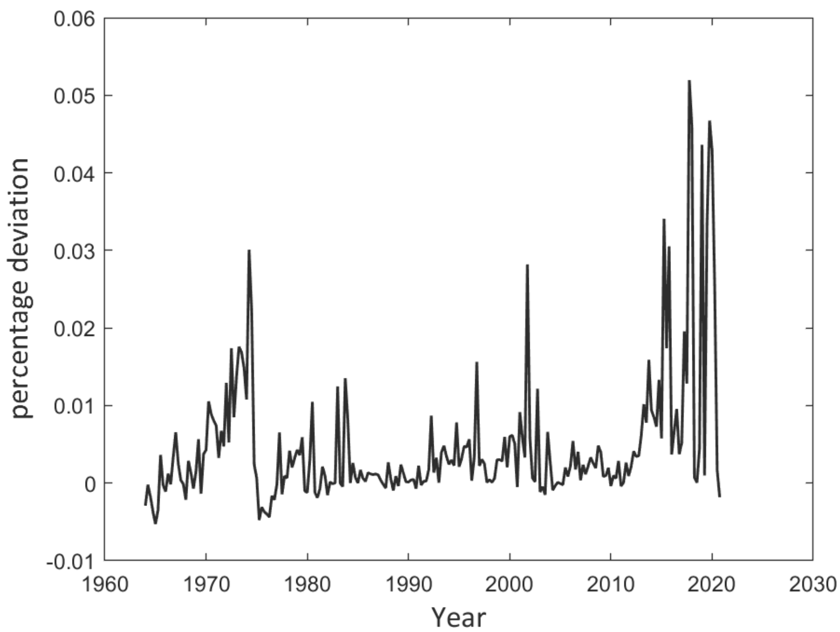
Notes: Impulse responses of output (Y) to government consumption (G) shock on impact across time. The date series on the x-Axis indicates end dates of the rolling window samples of size 40.

Fig. 3.4: IRF of Y to G shock at four quarters



Notes: Impulse responses of output (Y) to government consumption (G) shock at four quarters across time. The date series on the x-Axis indicates end dates of the rolling window samples of size 40.

Fig. 3.5: IRF of Y to G shock at 8 quarters



Notes: Impulse responses of output (Y) to government consumption (G) shock at eight quarters impact across time. The date series on the x-Axis indicates end dates of the rolling window samples of size 40.

3.11 Conclusion

In this chapter, I analyse the forecasting performance of DSGE models via the chapter 2 model for output, government consumption and debt forecasting through the Bayesian DSGE-VAR framework. The DSGE-VAR approach developed by Del Negro and Schorfheide (2004) aims to combine the forecasting performance of VARs with the structural and economic foundation of the DSGE model. The results in this chapter confirm previous results that the DSGE-VAR framework offers useful parameter shrinkage like other Bayesian priors and, at the same time, seems to aid forecasting performance for output forecasting. In addition to previous research, the DSGE-VAR approach can also effectively improve the forecasting performance of VARs for fiscal variables like government consumption and debt providing additional evidence for the methodology developed by Del Negro and Schorfheide (2004).

Based on this, it seems that the structural design of fiscal policy and its interactions with the economy, as postulated in the chapter 2 model and, by extension, other DSGE models can be useful for forecasting fiscal data over a purely uninformative prior as in standard VARs. For all data series, the economic intuition weaved into the fabric of DSGE models seems to be particularly useful at improving forecasting accuracy over longer horizons. Furthermore, VAR models of higher order can benefit greatly from DSGE priors, making them competitive with much smaller models. As shown in this chapter, higher-order frequentist VARs can struggle quite significantly with the small datasets common to macroeconomics. To circumvent this, the DSGE-VAR framework provides a complete and useful prior defined on VARS of arbitrary orders and the resulting parameter shrinkage proves effective in improving forecasting performance. However, there are exceptions to these results, and it seems there is some nuance to the benefit of DSGE models that may depend on sample selection, modelling framework and potentially other factors. In that, the results are similar to Gürkaynak, Kısacıkoğlu and Rossi (2014) in that there does not seem to be a single best methodology, but rather, some methodologies seem to work better in some scenarios and worse in others.

An additional contribution comes in the form of including a zero-lower bound model in the forecasting comparison. In comparison to the DSGE-VAR approach, the zero-lower bound approach faces a delicate trade-off between computational complexity and modelling advantages. In this analysis, the zero-lower bound model fares well for output forecasting but not better than comparatively simpler approaches. Specifically, in the financial crisis period, the performance of the ZLB modelling approach is not encouraging. For government consumption and debt forecasting, the ZLB approach seems to be more useful for rolling window estimations and less so for expanding windows. On a last note, on the ZLB approach, the estimation results are based on this application of the methodology, and it seems that some components are quite sensitive to the exact programming implementation. Therefore, the results here ought to be taken with a healthy dose of scepticism.

The last contribution comes as a by-product of the rolling-window estimations used to compare forecasting performance. The DSGE-VAR approach conveniently also generates posterior parameter estimates for the DSGE, which is identified by assumption. This allows for the impulse responses to be traced out at no significant computational burdens. This chapter explores how impulse responses vary across time for government consumption for prediction purposes. Based on the estimation here, estimates seem to be time-varying and, thus, depend on the estimation sample. Specifically, for samples that include the financial crisis data, this chapter shows that impulse response estimates are substantially higher than the time average.

3.12 References

Adolfson, M., Lindé, J. and Villani, M. (2007) “Forecasting performance of an open economy DSGE model,” Econometric Reviews, 26(2-4), pp. 289–328. Available at: https://doi.org/10.1080/07474930701220543.

Amisano, G. and Tristani, O. (2010) “Euro area inflation persistence in an estimated nonlinear DSGE model,” Journal of Economic Dynamics and Control, 34(10), pp. 1837–1858. Available at: https://doi.org/10.1016/j.jedc.2010.05.001.

Babecký, J., Franta, M. and Ryšánek, J. (2018) “Fiscal policy within the DSGE-VAR framework,” Economic Modelling, 75, pp. 23–37. Available at: https://doi.org/10.1016/j.econmod.2018.06.005.

Boehl, G. and Strobel, F. (2022) “Estimation of DSGE models with the effective lower bound,” SSRN Electronic Journal [Preprint]. Available at: https://doi.org/10.2139/ssrn.4138532.

Cai, M. et al. (2019) “DSGE forecasts of the lost recovery,” International Journal of Forecasting, 35(4), pp. 1770–1789. Available at: https://doi.org/10.1016/j.ijforecast.2018.12.001.

Chauvet, M. and Potter, S. (2013) “Forecasting output,” Handbook of Economic Forecasting, pp. 141–194. Available at: https://doi.org/10.1016/b978-0-444-53683-9.00003-7.

Christiano, L.J., Eichenbaum, M. and Evans, C.L. (2005) “Nominal rigidities and the dynamic effects of a shock to monetary policy,” Journal of Political Economy, 113(1), pp. 1–45. Available at: https://doi.org/10.1086/426038.

Christoffel, K., Coenen, G. and Warne, A. (2012) “Forecasting with DSGE models,” The Oxford Handbook of Economic Forecasting, pp. 89–128. Available at: https://doi.org/10.1093/oxfordhb/9780195398649.013.0005.

Del Negro, M. and Schorfheide, F. (2004) “Priors from general equilibrium models for vars,” International Economic Review, 45(2), pp. 643–673. Available at: https://doi.org/10.1111/j.1468-2354.2004.00139.x.

Del Negro, M. and Schorfheide, F. (2010) Bayesian Macroeconometrics, Northwestern University. Available at: https://faculty.wcas.northwestern.edu/lchrist/course/Korea_2016/del_negro_schorfheide.pdf (Accessed: April 7, 2023).

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), pp. 391–405. Available at: https://doi.org/10.1016/j.jeconom.2016.02.006.

Diebold, F.X., Schorfheide, F. and Shin, M. (2017) “Real-time forecast evaluation of DSGE models with Stochastic volatility,” Journal of Econometrics, 201(2), pp. 322–332. Available at: https://doi.org/10.1016/j.jeconom.2017.08.011.

Doan, T., Litterman, R. and Sims, C. (1984) “Forecasting and conditional projection using realistic prior distributions,” Econometric Reviews, 3(1), pp. 1–100. Available at: https://doi.org/10.1080/07474938408800053.

Edge, R.M. and Gürkaynak, R.S. (2011) How useful are estimated DSGE model forecasts, Board of Governors of the Federal Reserve System. Available at: https://www.federalreserve.gov/econres/feds/how-useful-are-estimated-dsge-model-forecasts.htm (Accessed: April 7, 2023).

Gelfer, S. (2019) “Data-rich DSGE model forecasts of the Great Recession and its recovery,” Review of Economic Dynamics, 32, pp. 18–41. Available at: https://doi.org/10.1016/j.red.2018.12.005.

Gelfer, S. (2021) “Evaluating the forecasting power of an open-economy DSGE model when estimated in a data-rich environment,” Journal of Economic Dynamics and Control, 129, p. 104177. Available at: https://doi.org/10.1016/j.jedc.2021.104177.

Ghent, A.C. (2009) “Comparing DSGE-Var forecasting models: How big are the differences?” Journal of Economic Dynamics and Control, 33(4), pp. 864–882. Available at: https://doi.org/10.1016/j.jedc.2008.10.004.

Giovannini, M., Pfeiffer, P. and Ratto, M. (2021) Efficient and robust inference of models with occasionally binding constraints, joint-research-centre. Available at: https://joint-research-centre.ec.europa.eu/document/download/08470a20-6992-4848-b40f-57970e6dcb8e_en (Accessed: April 7, 2023).

Guerrieri, L. and Iacoviello, M. (2015) “Occbin: A toolkit for solving dynamic models with occasionally binding constraints easily,” Journal of Monetary Economics, 70, pp. 22–38. Available at: https://doi.org/10.1016/j.jmoneco.2014.08.005.

Gupta, R. and Steinbach, R. (2013) “A DSGE-VAR model for forecasting key South African macroeconomic variables,” Economic Modelling, 33, pp. 19–33. Available at: https://doi.org/10.1016/j.econmod.2013.03.012.

Gürkaynak, R.S., Kısacıkoğlu, B. and Rossi, B. (2014) “DO DSGE models forecast more accurately out-of-sample than VAR models?,” Var Models in Macroeconomics - New Developments and Applications: Essays in Honor of Christopher A. Sims, pp. 27–79. Available at: https://doi.org/10.1108/s0731-905320130000031002.

Klein, P. (2000) “Using the generalized Schur form to solve a multivariate linear rational expectations model,” Journal of Economic Dynamics and Control, 24(10), pp. 1405–1423. Available at: https://doi.org/10.1016/s0165-1889(99)00045-7.

Kolasa, M. and Rubaszek, M. (2015) “Forecasting using DSGE models with financial frictions,” International Journal of Forecasting, 31(1), pp. 1–19. Available at: https://doi.org/10.1016/j.ijforecast.2014.05.001.

Leeper, E.M., Plante, M. and Traum, N. (2010) “Dynamics of fiscal financing in the United States,” Journal of Econometrics, 156(2), pp. 304–321. Available at: https://doi.org/10.1016/j.jeconom.2009.11.001.

Litterman, R. (1986) “Forecasting with bayesian vector autoregressions — five years of experience,” International Journal of Forecasting, 2(4), pp. 497–498. Available at: https://doi.org/10.1016/0169-2070(86)90100-7.

Nalban, V. (2018) “Forecasting with DSGE models: What frictions are important?” Economic Modelling, 68, pp. 190–204. Available at: https://doi.org/10.1016/j.econmod.2017.07.015.

Negro, M.D. and Schorfheide, F. (2013) “DSGE model-based forecasting” Handbook of Economic Forecasting, pp. 57–140. Available at: https://doi.org/10.1016/b978-0-444-53683-9.00002-5.

Pagan, A. (2003) Report on the modelling and forecasting at the Bank of England, www.bankofengland.co.uk. Available at: https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2003/report-on-modelling-and-forecasting-at-the-boe.pdf (Accessed: April 25, 2023).

Pfeiffer, J. (2018) A Guide to Specifying Observation Equations for the Estimation of DSGE Models. randomwalk.top. Available at: https://randomwalk.top/wp-content/uploads/2020/04/Pfeifer_2013_Observation_Equations.pdf (Accessed: April 7, 2023).

Pop, R.-E. (2017) “A small-scale DSGE-Var model for the Romanian economy,” Economic Modelling, 67, pp. 1–9. Available at: https://doi.org/10.1016/j.econmod.2016.07.011.

Smets, F. and Wouters, R. (2003) “An estimated dynamic stochastic general equilibrium model of the Euro Area,” Journal of the European Economic Association, 1(5), pp. 1123–1175. Available at: https://doi.org/10.1162/154247603770383415.

Smets, F. and Wouters, R. (2007) “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” American Economic Review, 97(3), pp. 586–606. Available at: https://doi.org/10.1257/aer.97.3.586.

Todd, R.M. (1984) “Improving economic forecasting with Bayesian vector autoregression,” Quarterly Review, 8(4). Available at: https://doi.org/10.21034/qr.843.

Wickens, M. (2014) “How useful are DSGE macroeconomic models for forecasting?,” Open Economies Review, 25(1), pp. 171–193. Available at: https://doi.org/10.1007/s11079-013-9304-6.

Appendix A

Appendix to Chapter 1

A.1 Data collection

The data set was constructed as close to Leeper, Plante and Traum (2010)'s original dataset as possible to allow for comparability. Data was collected from Q1 1947 to Q1 2019 (not including). For model estimation purposes only data from 1960 onwards is used to mirror Leeper, Plante and Traum (2010). This is relevant, because some of the series used by Leeper, Plante and Traum (2010) are not available on a quarterly basis before 1960. Additionally, this chapter does not rely on the constructed debt series as it turned out to be difficult to replicate (see debt series construction below).

For some data series, some concessions were made due to unavailability or other data issues and some table references have changed. This appendix gives detail when or what changes were made.

Most of the data was collected from the Bureau of Economic Analysis tables (BEA) and Federal Reserve data banks (FRED). See the following breakdown for more detail.

Household data

Consumption, C, is defined as the sum of personal consumption expenditure on nondurable goods (BEA table 1.1.5, Row 5) and services (BEA table 1.1.5, Row 6).

Investment, I, is specified as the sum of personal consumption expenditures on durables goods (BEA table 1.1.5, Row 4) and gross private domestic investment (BEA table 1.1.5, Row 7).

The labour observables variables, hours worked, is constructed the following way:

$$L = \frac{H * E}{100},$$

where H is defined as average weekly hours in the nonfarm business sector with index in Q1 1992 (Q1 1992=100) taken from the FRED data base (FRED table PRS85006023_NBD19920101). Additionally, E is civilian employment of 16 years or older in thousands of persons (FRED table CE160V) converted into an index with Q3 1992=100 as per Leeper.

Government

The consumption tax revenues, T^c, are the taxes on production and imports series (BEA table 3.2, Row 4) which includes both excise taxes and custom duties as per Leeper. Additionally, the state and local sales tax, T^c_s, is taken BEA tables 3.3 (Row 7).

Combining the previously described consumption related series, the consumption tax rate is defined in the following way:

$$\tau^c = \frac{T^c}{C - T^c \ - T^c_s}.$$

Both government capital tax income, T^k, and labour tax income, T^l, are generated by first creating tax rate series ($\tau^k$and τ^l, respectively) and then multiplying it with the respective tax bases. For the computation of both series, first, the average personal income tax rate as per Jones (2002) is calculated:

$$\tau^p = \frac{IT}{W + \frac{PRI}{2} + CI}.$$

Capital income, CI, is defined as the rental income (BEA table 1.12, Row 12), corporate profits (BEA table 1.12, Row 13) and interest income (BEA table 1.12, Row 18). Wage and salary accruals, W, is taken from BEA table 1.12 (Row 3) and proprietors income, PRI, is from BEA table 1.12 as well (row 12).

Then, the average capital income tax rate is defined as:

$$\tau^k = \frac{\tau^p CI + CT}{CI + PT}.$$

In this, CT are taxes on corporate income (BEA table 3.2, Row 8) and PT are property taxes (BEA table 3.3, Row 9).

The average labour income tax rates is the following:

$$\tau^l = \frac{\tau^p (W + PRI/2) + CSI}{EC + PRI/2}.$$

CSI are the contributions to government social insurance (BEA table 3.2, Row 11) and EC are compensations to employees (BEA tables 1.12, Row 2).

Then in the final step, T^k is defined as the multiplication of τ^k with CI and T^l as the multiplication τ^l with W as a measure of the respective tax bases.

Government expenditure, G, is defined as the sum of federal government gross investment (BEA table 3.2, Row 45), federal government consumption (BEA table 3.2, Row 25) and government net purchases of non-produced assets (BEA table 3.2, Row 47) minus government consumption of fixed capital (BEA table 3.2, Row 48).

Federal Government Transfers, Z, as the sum of federal subsidies (BEA table 3.2, Row 36), net current transfers (BEA table 3.2, row 28 minus row 19), net capital transfers (BEA table 3.2, row 46 minus row 42) minus the tax residual. The tax residual is defined as the sum of current tax receipts (BEA table 3.2, row 2), contributions for government social insurance (BEA table 3.2, row 10), income receipts on assets (BEA table 3.2, row 13) and the current surplus of government enterprises (BEA table 3.2, row 33) minus the sum of the previously described tax incomes ($T = T^c + T^k + T^l$).

The government debt series, B, could not be replicated, as the constructed series did not perform well. Instead, this chapter uses the Dallas fed. Market Value of U.S. federal debt series, which is transformed to quarterly data via simple averaging.

Construction of model observables

The previous section explored the data collection for the model observables (consumption (C), investment (I), hours worked (L), government spending (G), labour tax revenues (T^l), capital tax revenues (T^k), consumption tax revenues (T^c), government debt (B) and transfers (Z). In the next step, the nominal data is first converted to real data via deflating with the GDP deflator of personal consumption expenditures obtained from the BEA tables 1.1.4 (Row 2). Then the following transformation is made using a population index constructed from civilian noninstitutional population, years 16 and older, (FRED table: CNP16OV) with Q3 1992 as the base year (Q3 1992=1):

$$X = ln\left(\frac{x}{Popindex}\right) * 100,$$

where x indicates the original deflated series. Then, in a final step, the X series are detrended and demeaned using a linear trend specification to obtain the model observables. This detrending is done very every subsample estimation as well to ensure mean zero processes.

Generally, the constructed variables perform very well as they are close to the Leeper, Plante and Traum (2010) data set with correlations coefficients in the high nineties:

Table A.1: Correlation comparison with Leeper, Plante and Traum (2010) data

Variables	replication data correlation
I	0.985
C	0.988
L	0.999
T^c	0.998
T^l	0.982
Z	0.978
T^k	0.976
G	0.980
B	0.947

Notes: The correlation measure is the standard correlation coefficient.

A.2 RWMH estimation

For the replication of the estimation results presented in Leeper, Plante and Traum (2010), I utilize their estimation procedures which are equivalent to the Random Walk Metropolis Hasting (RWMH) algorithm presented by Herbst and Schorfheide (2016). Metropolis-Hastings samplers construct a Markov Chain, which asymptotically converges to the posterior distribution. Metropolis-Hastings samplers rely on proposal kernels to transition between states. Ideally, the choice of proposal kernel is tailored to the actual posterior to ensure efficient candidates are sampled. In the RWMH algorithm used in Leeper, Plante and Traum (2010), the proposal distribution is a normal distribution centred at the previous draw. The covariance matrix of the kernel is set to the inverse of the hessian obtained at the mode of the posterior distribution. The mode is estimated using a quasi-Newton method with a BFGS update of the hessian as part of csminwel developed by Christoper Sims. The simulation is then initialized using a draw from the prior, and the Metropolis-Hastings sampler generates 5,000,000 draws to ensure convergence. In addition, the first 250,000 draws are discarded as burn-in to avoid including draws that are potentially very far away from high-density regions of the posterior. Every 200th draw is thinned to create samples that are as uncorrelated as possible.

A.3 SMC estimation

This chapter relies on the Sequential Monte Carlo sampler's code based on Herbst and Schorfheide (2016) for the estimation. At their core, Sequential Monte Carlo samplers derive a sequence of Importance approximations that start at the prior distribution and end at the posterior. Given a good proposal distribution, the advantage of importance sampling is that it can generate precise approximations of the target density. This holds even if the density has non-gaussian properties like multimodality and unusually heavy weight on the tails. Typically, choosing good proposal distributions is essential to the success of importance sampling. The proposal distribution problem is solved in SMC samplers elegantly by generating a series of tempered importance approximations. In this, starting at the prior, the previous approximation is used as a proposal for the current tempered density. If the tempered densities are never too far away from each other, the particles can sufficiently adjust. The approximation will then be accurate. In the following, the mechanism is shortly presented in a low-detail form (for more detail, see Herbst and Schorfheide (2016) or (2014)):

Draw N initial particles from the prior $\theta_i \sim p(\theta)$ with equal importance weights

1. For $i = 1, \ldots, N_\phi$ // sequentially for all iterations
   1.1. Correction: Adjust particle weights to the current temperature and normalise the weights
   1.2. Selection: Resample the particles if the importance distribution is very uneven.
   1.3. Mutation: Let all particles adapt to the current temperature via Metropolis-type steps
2. Generate final importance sampling approximation

For each tempered distribution in the series from 1 to N_{ϕ}, the sampler first adjusts the particles to the current distribution by adjusting the weights and afterwards normalizes the weights to obtain the current approximation. In case the distribution is very uneven, the particles are then resampled. Ideally, less important particles are sampled out while particles with higher density are used to repopulate the sample. Typically, this resampling step reduces the diversity of the particle system. To combat this, the mutation step aims at reintroducing this diversity. The individual particles are adapted to the current density via Metropolis-Hastings steps to increase diversity. However, it's important to note that this step is not necessary to ensure that the particles are drawn from the current tempered distribution and only seeks to reintroduce variation.

Because DSGE models typically are quite large dimensionally and tend to create difficult posteriors, the selection and mutation steps must be carefully chosen and well-tuned to keep the particles alive and healthy during the estimation process. For the selection step, the residuals are resampled using systematic resampling. For this, the script published by Herbst and Schorfheide (2016) is used. The Mutation step is equally important and ensures that the importance approximation is proper. It does so by 'jittering' the particles at each temperature. Ideally, this allows the particles to adapt properly to the current temperature and the tempered posterior density. What happens if the mutation step malfunctions is that relatively little new information is introduced, and the particles can become stuck. In combination with the resampling step, this means that over time the particle distribution collapses, causing sample impoverishment. Following Herbst and Schorfheide (2016), this chapter implements random blocking in the mutation step with a proposal generated from a mixture density. The two additions improve the mutation step in two directions. Firstly, random blocking reduces the number of parameters that are mutated at the same time by breaking down the parameter vector into smaller blocks. Consequently, the mutation probability increases, and this essentially allows the particles to move quicker through the parameter space. Secondly, the mixture proposal works as follows for a specific block:

$$\nu_{n,b}|(\theta^i_{n,b}, \theta^i_{n,-b}, \theta^*_{n,b}, \Sigma^*_{n,b}) \sim wN(\theta^i_{n,b}, c^2\Sigma^*_{n,b}) + \frac{1-w}{2} N\left(\theta^i_{n,b}, c^2 diag(\Sigma^*_{n,b})\right)$$
$$+ \frac{1-w}{2} N(\theta^*_{n,b}, c^2\Sigma^*_{n,b}).$$

In addition to the standard proposal with weight w, it also has an independent proposal from the mean and a proposal with a diagonal covariance matrix. The former allows particles to move independently from the current position of the particle, and the latter eases on the co-movement of parameters locally. Together, the mixture proposal and blocking improve the movement of particles.

The SMC sampler requires an econometrician to determine a set of tuning parameters for the estimation to work optimally. The tuning parameters (the number of blocks, N_b, the tempering parameter, λ, the number of temperatures, N_ϕ, and the mixture weight, w) are set identically to Herbst and Schorfheide (2016). In the replication section, N is set to 6000, consistent with the original estimation. In the following sections, the number of particles is increased to 20000. The advantage is that the accuracy of the approximation of the posterior is increasing in the number of particles, and it allows the sampler to better maintain a multimodal posterior. The tuning parameter settings are summarized in Table A.2 below.

Table A.2: Tuning Parameters for the SMC sampling

parameter	value
N	6000/20000
N_b	3
λ	4
N_ϕ	500
w	0.9

Notes: N corresponds to the number of particles. N_b is the number of blocks. λ defines the curvature of the tempering schedule and over the number of iterations, N_ϕ. w is the weight of the mixture distribution.

A.4 HPD intervals

As shown in Herbst and Schorfheide (2016), non-gaussian and multimodal posterior distributions are expected in this chapter's estimations. Here, I address the topic of posterior uncertainty by using higher posterior density intervals and regions, as in Chen and Shao (1999) and Chen et al. (2000).

The interesting problem with multimodal parameter distributions is that the highest posterior density (HPD) intervals tend to overestimate the uncertainty in the object. To illustrate this, an HPD interval for a multimodal distribution with a low-density area between the modes will typically include the low-density area as, by definition, an interval may not include disjointed regions. As such, the distribution of interest will appear more diffuse than it actually is. This non-normality or multimodality can materialise itself not only in posterior parameter distributions but equally in other posterior objects like impulse responses where uncertainty is critical for policymakers. The way to solve this problem is to search for the highest posterior regions as opposed to intervals for multimodal distributions. If the highest posterior region can be disjointed, then the regions can adjust to omit the low-density area between modes. Consequently, depending on the distribution, HPD regions may show a different picture of the uncertainty than HPD intervals.

To address this, I implement the approaches proposed in Chen and Shao (1999) and Chen et al. (2000) to evaluate both empirical unimodal intervals and bimodal density regions. The mechanism works by generating a sample from the density and sorting the sample from the smallest to the largest value. For the unimodal interval, a simple minimisation problem is solved that minimises the distance between the j^{th} and $(j+(1-a)n)^{th}$ parameter or object interest values to obtain the $100*(1-\mathrm{a})\%$ HPD interval, where n is the sample size. Under unimodality of the density and uniqueness of the solution to the minimisation, this interval is shown to converge to the true interval in the limit. Formally, the minimisation problem is as follows for posterior parameter distributions. In the first step, a sample of size n is obtained from the posterior distribution of interest. In the second step, to pre-process, all elements of the sample are sorted in ascending order:

$$\theta_1 \leq \theta_2 \leq \ldots \leq \theta_n.$$

Based on the sorted sample, the following minimization problem is solved:

$$R_{j^*}(n) = \theta_{j^*+(1-a)n} - \theta_{j^*} = \min_{1\leq j\leq an} \theta_{j+(1-a)n} - \theta_j.$$

The problem minimizes the distance between the start and endpoint of the ordered sample subject to any sample, including $100*(1-\mathrm{a})\%$ of the sample observations. To ensure that the starting point can always generate a sample that can have $100*(1-\mathrm{a})\%$ of the sampled observations, the starting index is restricted to be weakly larger than one but weakly smaller than an. The HPD regions are derived from a very similar optimisation problem. This problem is restricted to two regions such that they correspond to a bimodal distribution. The problem is as follows:

$$R_{j^*,i^*,m^*}(n) = \min_{i,j,m}(\theta_{i+m} - \theta_i) + (\theta_{j+an-m} - \theta_j) \quad s.t. \quad 0 \leq m \leq (1-a)n \ , 0 \leq i$$
$$\leq n - n(1-a), \quad i + m \leq j \leq n - [(1-a)n - m].$$

This problem minimizes the sum of the sizes of the two intervals. Just like the unimodal problem, the bimodal problem is restricted in several ways. Firstly, the starting point of the first interval is set identically to the unimodal setting. The length of the first interval can include at most $100 * (1 - \mathrm{a})\%$ of the observations. The starting point of the second interval must be after or at the end of the first interval but needs to be set to ensure that both intervals together can include $100 * (1 - \mathrm{a})\%$ of the observations.

Both approaches are implemented in MATLAB using the mixed integer programming toolbox. The posterior intervals or regions of impulse responses estimated in this chapter are based on samples of size 1000, which is the same sample size Dynare uses by default. Parameter HPD intervals are evaluated using the full posterior.

A.5 Posterior estimates of non-fiscal parameters

The table below presents estimation results for non-fiscal rule parameters for the three variations of the Leeper, Plante and Traum (2010) model used in the main results section. That includes the standard (ind., sym., ind.) specification and the alternatives with (spending first, sym., ind.) and (taxation first, sym., ind.). Overall, the estimates appear stable across the specification and any deviations are easily contained in the posterior intervals of the original Leeper, Plante and Traum (2010) specification.

Table A.3: Parameter estimates of non-fiscal parameters

Param.	(spending first, sym., ind.) mean (standard dev.)	(taxation first, sym., ind.) mean (standard dev.)	(ind., sym., ind.) mean (standard dev.)
γ	2.07 (0.40)	2.01 (0.35)	2.06 (0.40)
κ	2.07 (0.45)	1.95 (0.40)	2.04 (0.44)
h	0.58 (0.07)	0.58 (0.07)	0.58 (0.07)
s''	5.58 (0.26)	5.59 (0.26)	5.53 (0.26)
δ_2	0.70 (0.29)	0.73 (0.30)	0.78 (0.36)
ρ_a	0.97 (0.01)	0.97 (0.01)	0.97 (0.01)
ρ_b	0.49 (0.03)	0.49 (0.03)	0.49 (0.04)
ρ_l	0.97 (0.01)	0.97 (0.01)	0.97 (0.01)
ρ_i	0.45 (0.07)	0.45 (0.07)	0.45 (0.07)
σ_a	0.63 (0.03)	0.63 (0.03)	0.63 (0.03)
σ_b	9.15 (0.56)	9.22 (0.58)	9.17 (0.59)
σ_l	2.75 (0.44)	2.63 (0.39)	2.72 (0.43)
σ_i	5.27 (0.45)	5.26 (0.45)	5.28 (0.48)

Notes: For a detailed description of the parameters and their purpose, see the model description in the main body of the chapter.

Appendix B

Appendix to Chapter 2

B.1 Second-order pruned system

For the estimations in this chapter, I rely on the canonical pruned second-order system as in Andreasen, Andreasen, Fernández-Villaverde and Rubio-Ramírez (2017). The key idea of pruning comes from the shock transition in non-linear models. In linear models, no shock can push the linear model of a stable path. In non-linear modelling, this is not necessarily the case, even if the linear model is stable. The instability is introduced via the inclusion of the higher terms of the approximation. In order to ensure that the data generating process is stable, the system is pruned. The pruned second-order system can be summarized as follows:

$$x_{t+1}^L = H_x x_t^L + \sigma J v_{t+1}, \qquad v_{t+1} \sim N(0, I),$$

$$x_{t+1}^Q = 0.5 * h_{\sigma\sigma}\sigma^2 + H_x x_t^Q + 0.5 * H_{xx}(x_t^L \bigotimes x_t^L) + \sigma J v_{t+1}, \qquad v_{t+1} \sim N(0, I),$$

$$y_{t+1} = 0.5 * g_{\sigma\sigma} + G_x x_{t+1}^Q + 0.5 * G_{xx}(x_{t+1}^L \bigotimes x_{t+1}^L).$$

Working with the pruned system may seem like an approximation of sorts. In practice, it is very convenient to prune unstable paths. In fact, most particle filter applications rely on it, and so does the conditional particle filter. In its basic form, the second-order DSGE system is non-linear in its states. However, the state vector can be augmented to linearize the system without additional assumptions. Using the state vector $z_t = [x_t^L{}', x_t^Q{}',\ (x_t^L \bigotimes x_t^L)']$ the system can be rewritten as such:

$$z_{t+1} = c_2 + A_2 z_t + B_2 \zeta_{t+1},$$

where the system can be stated in matrix form as:

$$\begin{bmatrix} x_{t+1}^L \\ x_{t+1}^Q \\ (x_{t+1}^L \bigotimes x_{t+1}^L) \end{bmatrix}$$
$$= \begin{bmatrix} 0 \\ 0.5 * h_{\sigma\sigma} \\ \sigma^2 (J \bigotimes J) vec\big(I_{n_v}\big) \end{bmatrix} + \begin{bmatrix} H_x & 0 & 0 \\ 0 & H_x & H_{xx} \\ 0 & 0 & (H_x \bigotimes H_x) \end{bmatrix} \begin{bmatrix} x_t^L \\ x_t^Q \\ (x_t^L \bigotimes x_t^L) \end{bmatrix}$$
$$+ \begin{bmatrix} \sigma J & 0 & 0 & 0 \\ \sigma J & 0 & 0 & 0 \\ 0 & \sigma^2 (J \bigotimes J) & \sigma^2 (J \bigotimes H_x) & \sigma^2 (H_x \bigotimes J) \end{bmatrix} \begin{bmatrix} v_{t+1} \\ (v_{t+1} \bigotimes v_{t+1}) - vec\big(I_{n_v}\big) \\ (v_{t+1} \bigotimes x_t^L) \\ (x_t^L \bigotimes v_{t+1}) \end{bmatrix},$$

In this augmented form, z_{t+1} depends linearly on z_t and ζ_{t+1}. ζ_{t+1} constitutes an uncorrelated, mean zero process with a finite covariance matrix under standard assumption. The covariance matrix of ζ_{t+1} can be constructed exactly based on the unconditional covariance matrix of the linear states. The measurement equation can be rewritten as follows:

$$y_{t+1} = d_2 + C_2 z_{t+1},$$

$$y_{t+1} = [0.5 * g_{\sigma\sigma}] + [0 \quad G_x \quad 0.5 * G_{xx}] \begin{bmatrix} x_{t+1}^L \\ x_{t+1}^Q \\ (x_{t+1}^L \bigotimes x_{t+1}^L) \end{bmatrix}.$$

Measurement errors for y_{t+1} can be added on demand as in any DSGE. Finally, while the new state-space system for $z_{t+1}$is of the canonical linear form, it is technically not gaussian. This is because ζ_{t+1} depends on higher-order products of v_{t+1}.

B.2 Code implementation detail

Solving and estimating higher-order DSGE models is a prohibitively time-consuming process that can require substantial code development and expensive hardware to become feasible. In the section below, I provide a detailed breakdown of the methods employed in this chapter, and hopefully, the review is useful to others implementing similar projects. As mentioned before, this work is heavily influenced by Gomme and Klein (2011), Schmitt-Grohe and Uribe (2004), Herbst and Schorfheide (2016), Buchholz, Chopis and Jacob (2021) and neural network applications.

To start off, for this project, MATLAB was the choice of programming language as it offers several advantages other languages do not currently offer for economics to the same extent. MATLAB is heavily used in macroeconomics for DSGE models, and a lot of important programs are freely available (e.g. solab.m/solab2.m for model solving developed as a companion to Gomme and Klein (2011)). While MATLAB is generally thought of as a low-performative language, it has made significant strides during the last 15 years. With the introduction of the Linear Algebra Package (LAPACK) library and Basic Linear Algebra Subprograms (BLAS) to MATLAB in the earlier 2000s, MATLAB itself has become performative and provides a great compromise between performance and speed of implementation. While implementation directly in C or Fortran ought to be faster, the implementation also becomes more time-consuming. However, to fully make use of the advancement, a requirement is that code is developed centred around the idea of utilizing the optimized libraries whenever suitable (vectorizing, utilizing compiled functions wherever possible, etc.).

For the code implementation, one component of particular importance is being able to adapt model files quickly and conveniently. Writing DSGE model files and the following debugging is a tricky and time-consuming process that requires frequent rewrites of sections and model matrices. At the moment, Dynare offers the best practice in the way it approaches model files. Dynare provides an incredibly convenient framework for writing models. The user can write the model equations directly using a convenient time-shift notation into a text file. Dynare then parses this file and creates estimation-relevant objects on the fly. The user does not have to manually supply any further objects like Jacobians, Hessians or linearized versions of the model. Adapting or changing a model is comparatively straightforward as only the model file has to be adjusted and requires no further input from the user. While I initially explored using Dynare, my experience was that it is based on a comparatively close-knitted ecosystem, and that makes it more difficult to, for example, replace estimation procedures or time series filters with non-Dynare options. Therefore, I re-engineered a basic version based on the Dynare philosophy for model files. I rely on the symbolic toolbox MATLAB provides, similar to Schmitt-Grohe and Uribe (2004). Symbolic variables are a data type, and the key idea is that it tells MATLAB that any calculations using these variables must be performed analytically. However, these symbolic variables can be evaluated as numerical variables. This provides significant freedom and convenience in their application. Further, MATLAB supports a wide range of functions for symbolic variables making them an ideal choice for quick and convenient development of model files. For the code implementation, the main model file that includes the DSGE equations is written entirely using symbolic variables, and the equations are based on the original non-linear system of equations.

Based on the symbolic variable implementation, several advantages come into play. Changing variables becomes much easier in this format. Instead of rewriting the system manually, the user only has to define a relation between the old and new variables. This relation can then be used to substitute the old variables out. As in Schmitt-Grohe and Uribe (2004), using the inbuilt symbolic functions allows the evaluation of estimation relevant objects for a specific DSGE model.

Crucially, the Jacobian and Hessian of the model must be evaluated for every single likelihood evaluation. While the analytic form of both matrices does not change, numerical evaluations vary. I evaluate the Jacobian and Hessian analytically once, in the beginning, using the $jacobian(f, v)$ and $hessian(f, v)$ functions based on the symbolic structural parameter vector. Next, the two objects are printed to a dynamically generated MATLAB function. During the simulation, the generated MATLAB function can then be used to evaluate the Jacobian and Hessian numerically using a numeric parameter vector. It requires no further use of the symbolic variables. This means that the evaluation only relies on a vector-valued matrix construction which is significantly more performative and convenient than other methods like numerical differentiation. Coming back to adaptability, model changes in the model files trickle down to this part. After making adjustments to the model files, a new file for evaluating the Jacobian/Hessian of the model can be generated without further adjustments to the code. Consequently, this type of implementation improves on the adaptability to model changes and adjustments in specification significantly while also being quite performative.

Convenience and adaptability aside, the crucial component of this application is performance. Non-linear estimations are much more time-consuming than linear estimations because they require additional complex calculations: model solving and likelihood construction.

Model solving has two main components that require significant time. The first component is generating the inputs for the model solver (i.e., the Jacobian and Hessian of the model evaluated at the steady state). As previously mentioned, I rely on printed analytical files for these objects, which can be easily and quickly evaluated. This essentially removes the inputs as a significant computation time cost factor. The second component is the solution strategy itself. For this, I rely on the alternative DSGE solution method offered by Gomme and Klein (2011) based on a generalized Sylvester equation approach and their code. Their implementation relies on calling LAPACK functions to solve the generalized Sylvester equation and offers substantial computational gains over all other tested implementations. Furthermore, it can be easily implemented.

The other performance-critical aspect is filtering/likelihood evaluations. Filtering for non-linear models can be done using particle filters which use thousands of particles to approximate the likelihood. Particle filters require numerous forward iterations of the individual particles per likelihood evaluation of the solved model. Because the number of particles is in the thousands typically and estimations require thousands if not millions of likelihood evaluations, the time per forward iteration is important. The forward iteration of the predetermined variables for a specific particle, x_t, is defined in the equation below:

$$x_{t+1} = 0.5 * h_{\sigma\sigma} + H_x x_t + 0.5 * H_{xx}(x_t \bigotimes x_t) + \sigma J v_{t+1}, \qquad v_{t+1} \sim N(0, I).$$

Specifically, the operation $H_{xx}(x_t \bigotimes x_t)$ for a given $(x_t \bigotimes x_t)$ is costly and has a time complexity of $O(n_x^3)$ using Big-O notation where n_x is the number of predetermined states. For a given structural parameter vector, H_{xx} is a fixed matrix. Because that is the case, the second-order component, $H_{xx}(x_t \bigotimes x_t)$, does need to be done particle by particle in a sequential format, but can be vectorized to one operation.[22] As it can be summarized to one operation, one can make good use of the inbuilt LAPACK libraries or run the process on a GPU in parallel. However, this process comes with a caveat typically not encountered during normal applications. Forwards iterating the entire particle system in one go requires significant amounts of memory, especially if it is done in parallel for several structural parameter vectors at once.

Based on MATLAB (R2021b), in parallel applications, MATLAB works out of the CPU cache. The CPU cache is a type of memory that is much faster and located close to the CPU. Any calculation that can be done on the cache is significantly faster than calculations based on data in the RAM or hard drive. The problem is that cache memory is much more expensive and therefore limited to a few MB. There are different types of caches with different speeds and sizes on the standard computers to optimize performance. The L1 type is usually the fastest but smallest. Other caches may address this trade-off between size and speed differently. In the case of MATLAB, in parallel applications, once a dataset is larger than a given CPU cache, performance will degrade because of memory access. Therefore, while intuitively, the speed of the CPU itself seems to be of utmost importance, in this application, memory access is the second important variable.

[22] All current particles can be stacked as column vectors into a matrix, X_t^{system} of size n_x by the number of particles. In this case, the Kronecker product can be conveniently defined using the *reshape* and *repelem* function as $repelem(X_t^{system}, n_x, 1) .* repmat(X_t^{system} n_x, 1)$, where ".*" indicates element by element multiplication. The last step is to pre-multiply by H_{xx}.

While the way MATLAB handles arrays that are larger than the CPU cache is mostly black box, for this application, it turns out that breaking down the calculation into sub-sets and ensuring that all required arrays for each parallel sub-calculation fit into the cache is advantageous. Fortunately, the forward iteration of the particle system can be broken down easily by separating the particle system into new arrays. To decide on this breakdown in a semi-optimal way, I approximate the total storage needed by the particle system and the arrays required in operation. The memory requirement is then divided by the cache size, which delivers the number of sub-sets of particles. Finally, this number is rounded up for a safety margin to avoid approximation errors. The result is a semi-efficient number of sub-sets for the total calculation.

To access any performance advantages, I test computation times for likelihood evaluations. The test computer has an Intel® Core™ i9-10980XE CPU. This CPU has three caches, where the L1 has 1.1MB of storage, the L2 has 18.0MB, and the L3 has 24.8 MB. I set the number of particles to 10.000, as in the estimation. The result is a significant speed-up for any cache, but in this case, the L1 cache seems to provide an optimal trade-off between size and speed. The gains are even starker for less performative CPUs, like in a work laptop.

To showcase the performance gains, Table B.1 below shows likelihood evaluations times for naïve and optimized memory management for single and multithreaded computations for the most performative cache (L1) based on a sample of 100 evaluations:

Table B.1: Overview of performance gains across different specifications

Processor	Memory management	Time/likelihood eval.
Single	Naïve	1.54
Single	optimized	0.82
Multi	Naïve	1.07
Multi	optimized	0.57
GPU	optimized	0.65

Notes: Time is measured in seconds

On a single Core, the time reduction from optimizing memory management is equal to roughly -48%. This also holds true for multithreaded calculations using parpool. Multithreaded calculations allow for an orthogonal performance gain to memory management, bringing the total gain up to 62% or 0.57s per likelihood evaluation. For Metropolis-Hastings samplers, this decrease scales linearly to estimation time as estimation time is roughly $n_{MHeval} * t_{likelihood\ eval.}$. For the sequential Monte Carlo estimation, I developed an alternative version in which the forward iteration is computed on the GPU which are already heavily employed in other fields that require large array operations (e.g. neural networks). Sequential Monte Carlo techniques (SMC) heavily rely on CPU parallelization. This can be a significant computational advantage over MCMC techniques. If the time per likelihood evaluation and the number of likelihood evaluations is fixed, then the SMC can be up to n_{Cores} times faster.[23] However, due to its parallel nature, CPU resources are typically not available for multithreading of the likelihood evaluation. The result is that the time/likelihood may be significantly slower if one purely relies on CPU calculations. The advantage of using a GPU for sequential Monte Carlo estimation is that it brings in new resources that can be split. Equally, GPUs have much larger memories, and typically, no cache-like constraints are experienced for the size of arrays in this problem. On a test basis for non-sequential likelihood evaluations, the time per likelihood evaluation for a GPU (0.65s) is comparable to that of the multithreaded memory-optimized implementation (0.57). As a last note, as the design of MATLAB memory access is mostly black box, this implementation need not always provide advantages, especially in different versions of MATLAB.

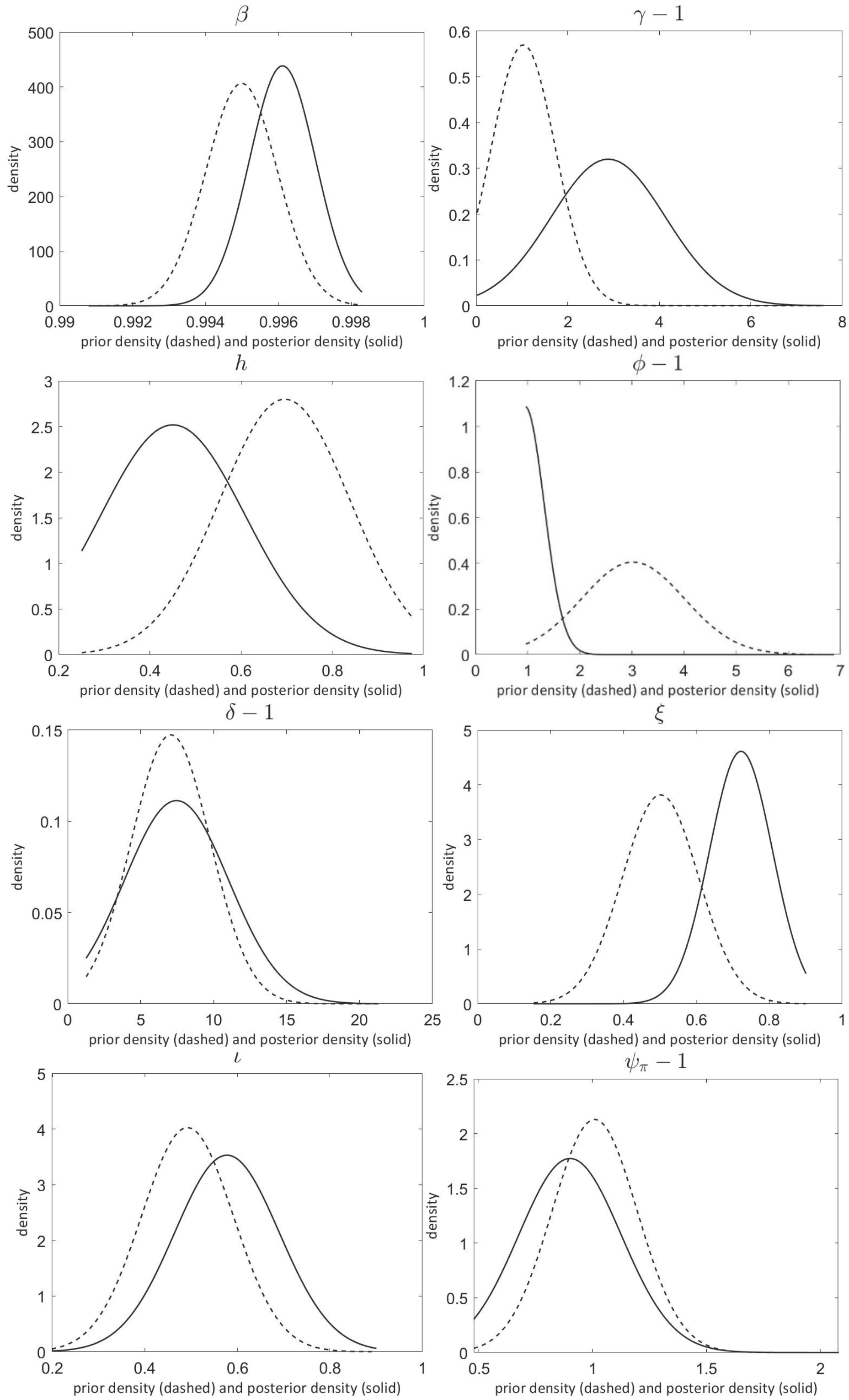
For this chapter, I am really grateful to Amisano and Tristani (2010) for providing me with their code base, which allowed me to double-check my work and improve upon it.

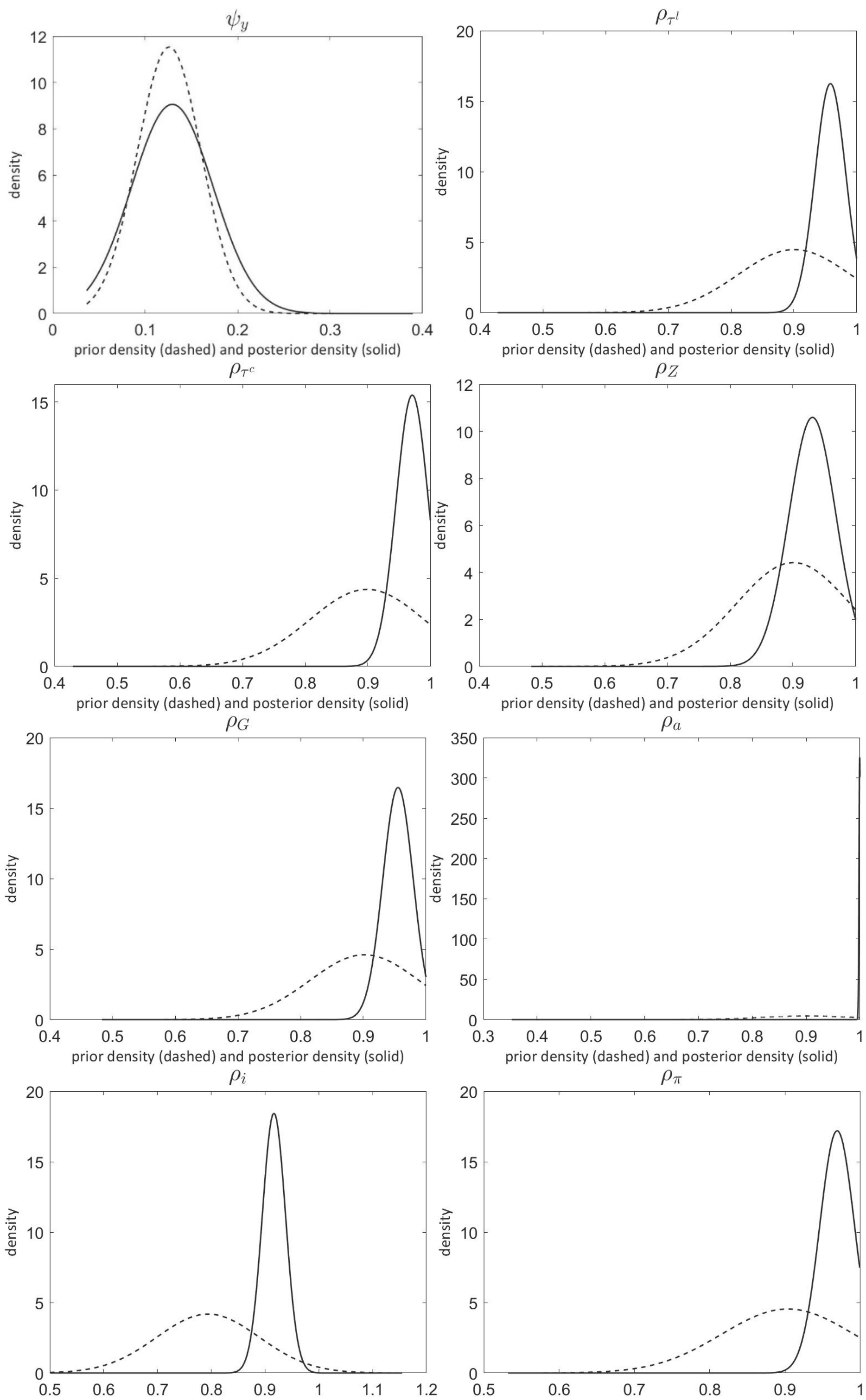
[23] The actual speed up depends on the exact algorithm settings (comparing reasonable algorithm settings vs forcing equal numbers of likelihood evaluations) but is substantial in any case. For an estimation using the Metropolis Hastings algorithm using a quite standard 5,000,000 likelihood evaluations as in Leeper, Plante and Traum (2010) estimation time may be as long as 89 (47, 32) days based on the single, naïve ((single, optimized), (Multi, optimized)) implementation. The run time for the SMC algorithm employed ended up being 5 days implying a reduction of 94% (89%, 85%). Comparing based on an equal number of likelihood evaluations, it would imply savings of 85% (72%, 60%).

B.3 Posterior density plots

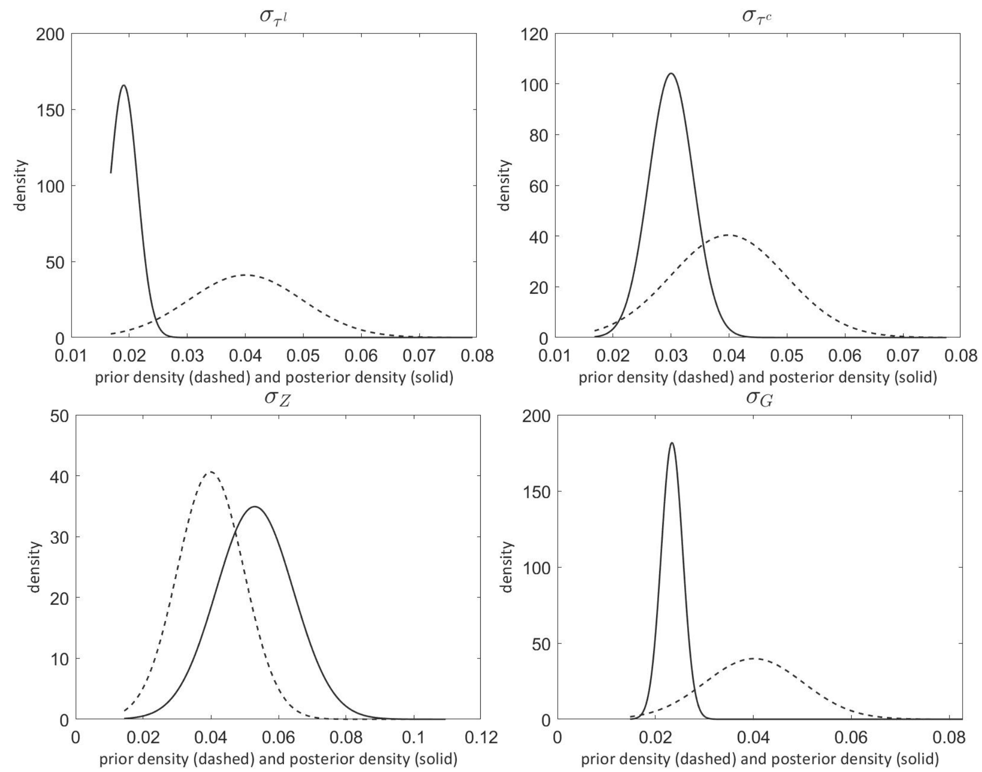
B.3.1 Core model parameters

Fig. B.1: Posterior density graphs for core model parameters





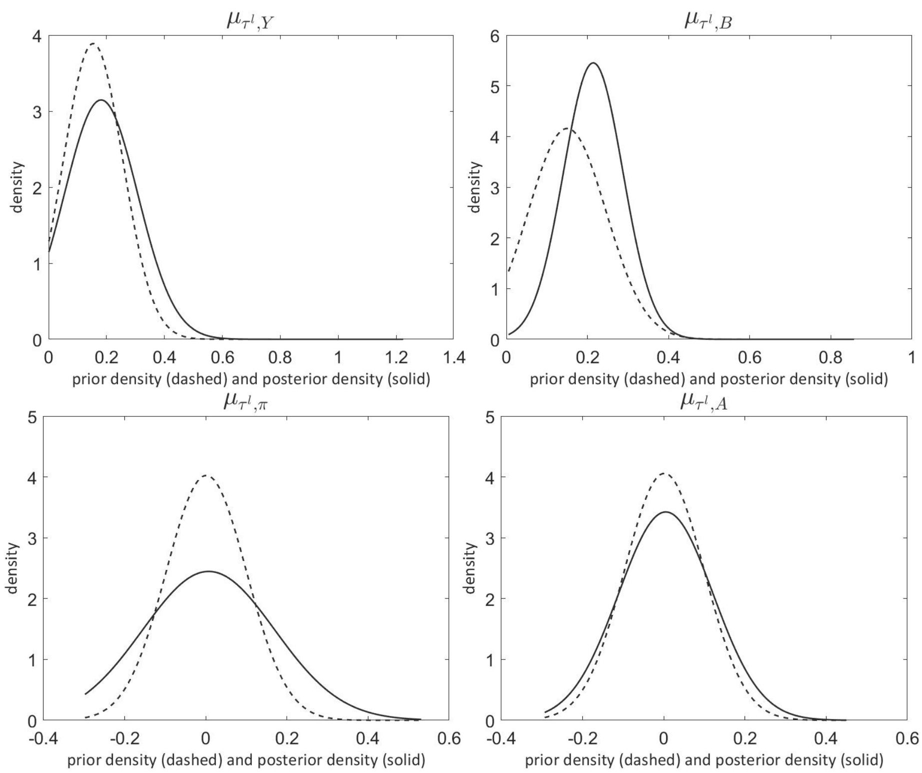
ψ_y
ρ_{τ^l}
ρ_{τ^c}
ρ_Z
ρ_G
ρ_a
ρ_i
ρ_π
density
prior density (dashed) and posterior density (solid)

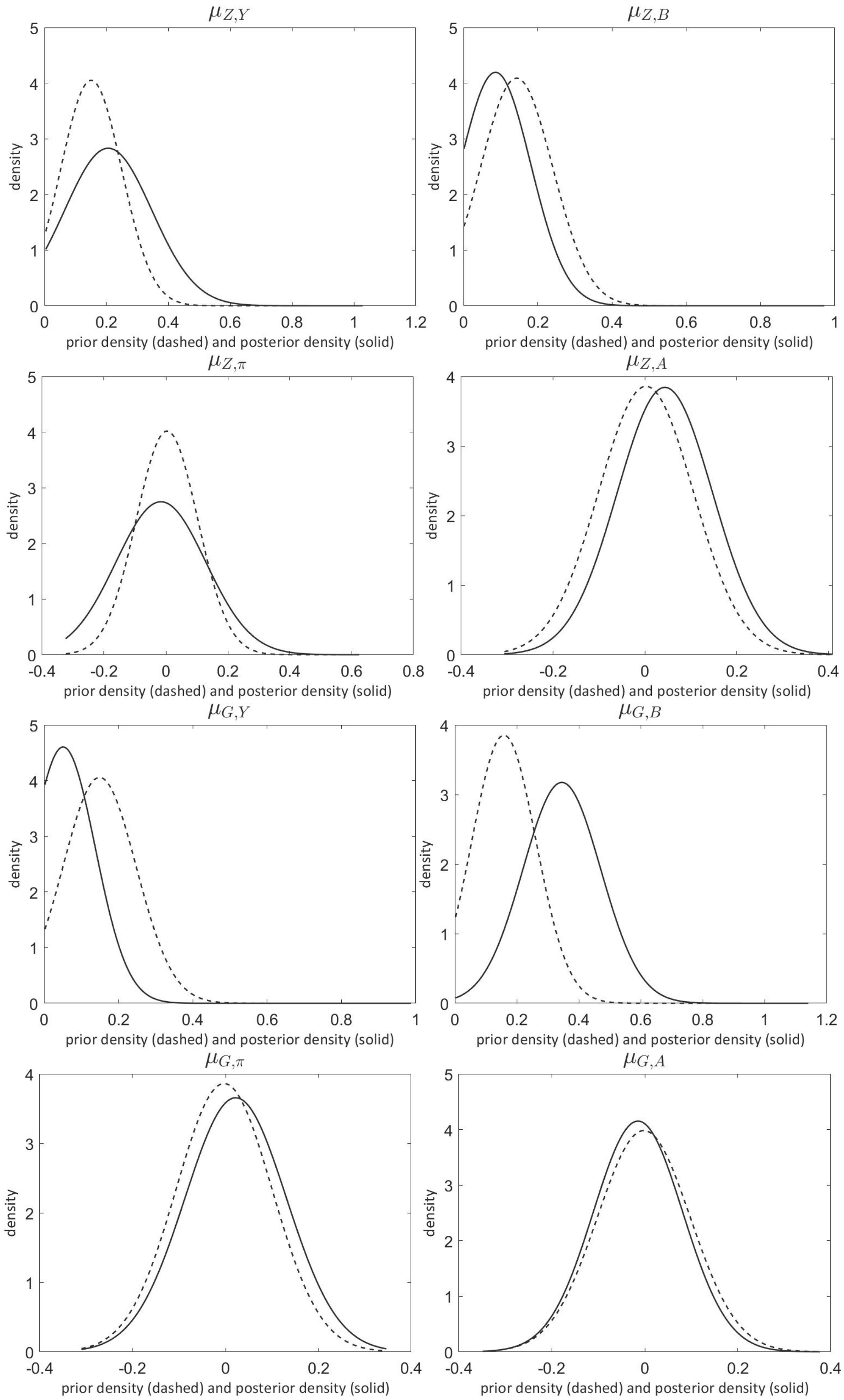


Notes: Posterior and prior density plots for core model parameters. Dashed line corresponds to prior density and solid line to posterior density.

B.3.2 Linear rule parameters

Fig. B.2: Posterior density graphs for linear rule parameters

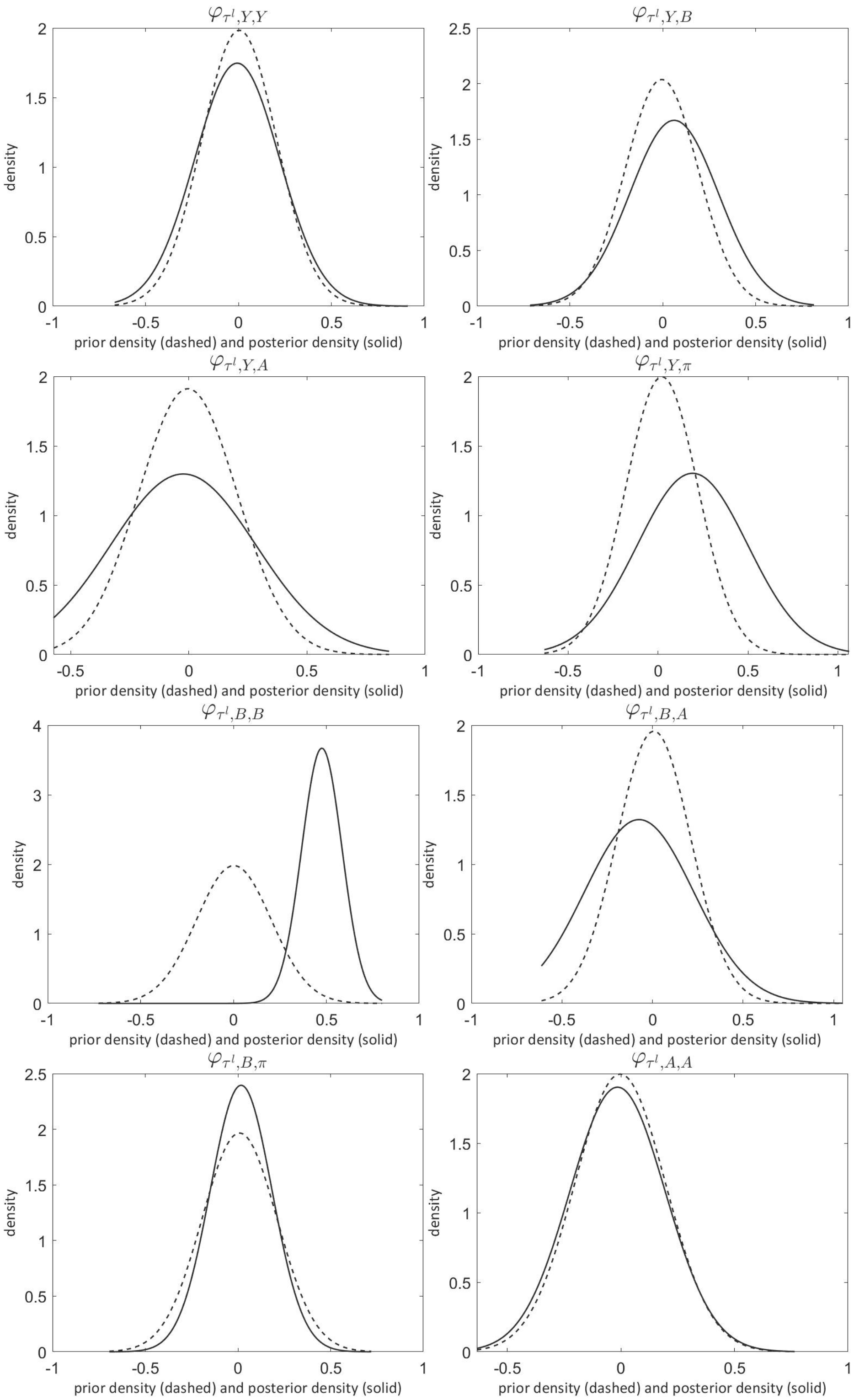


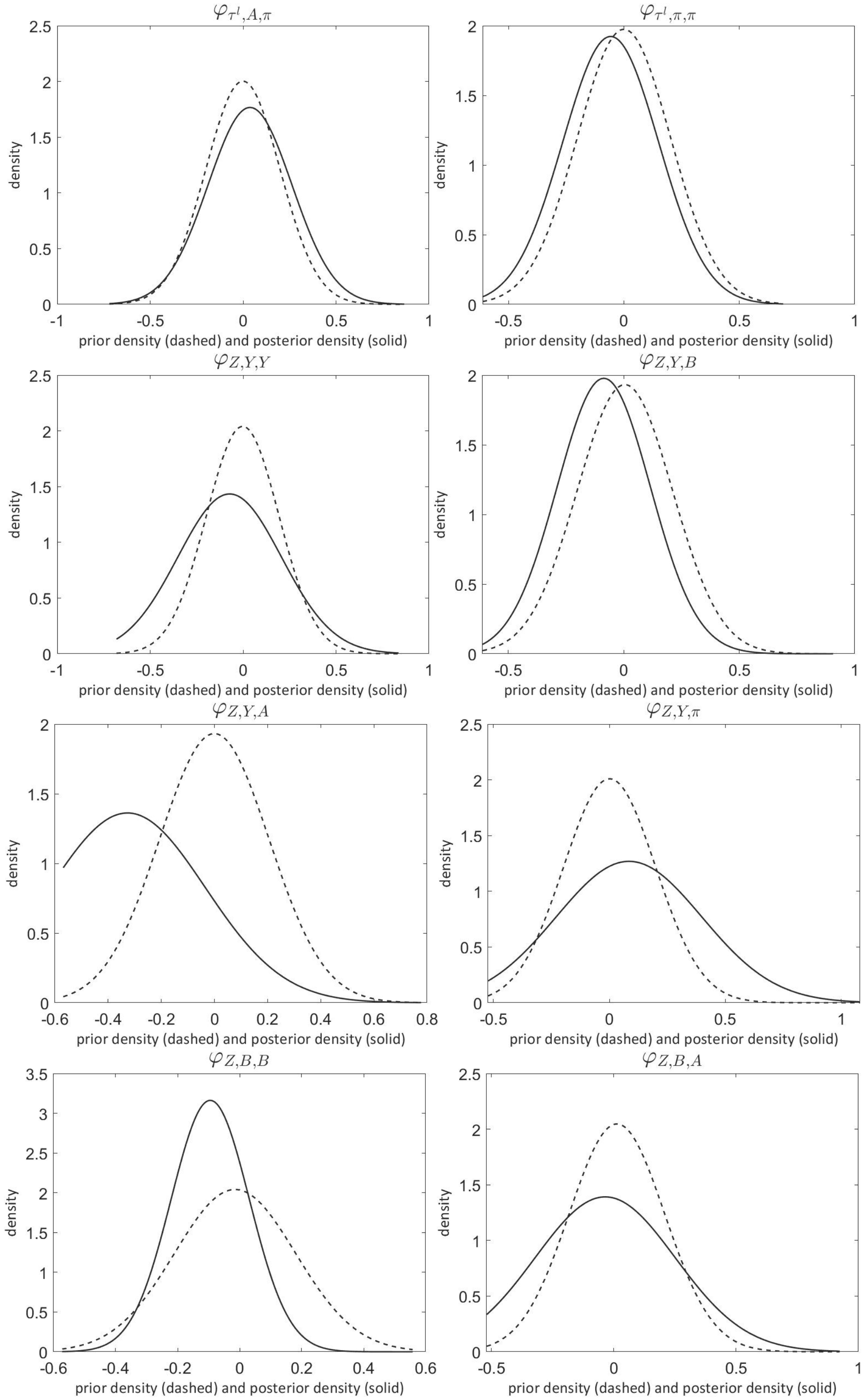


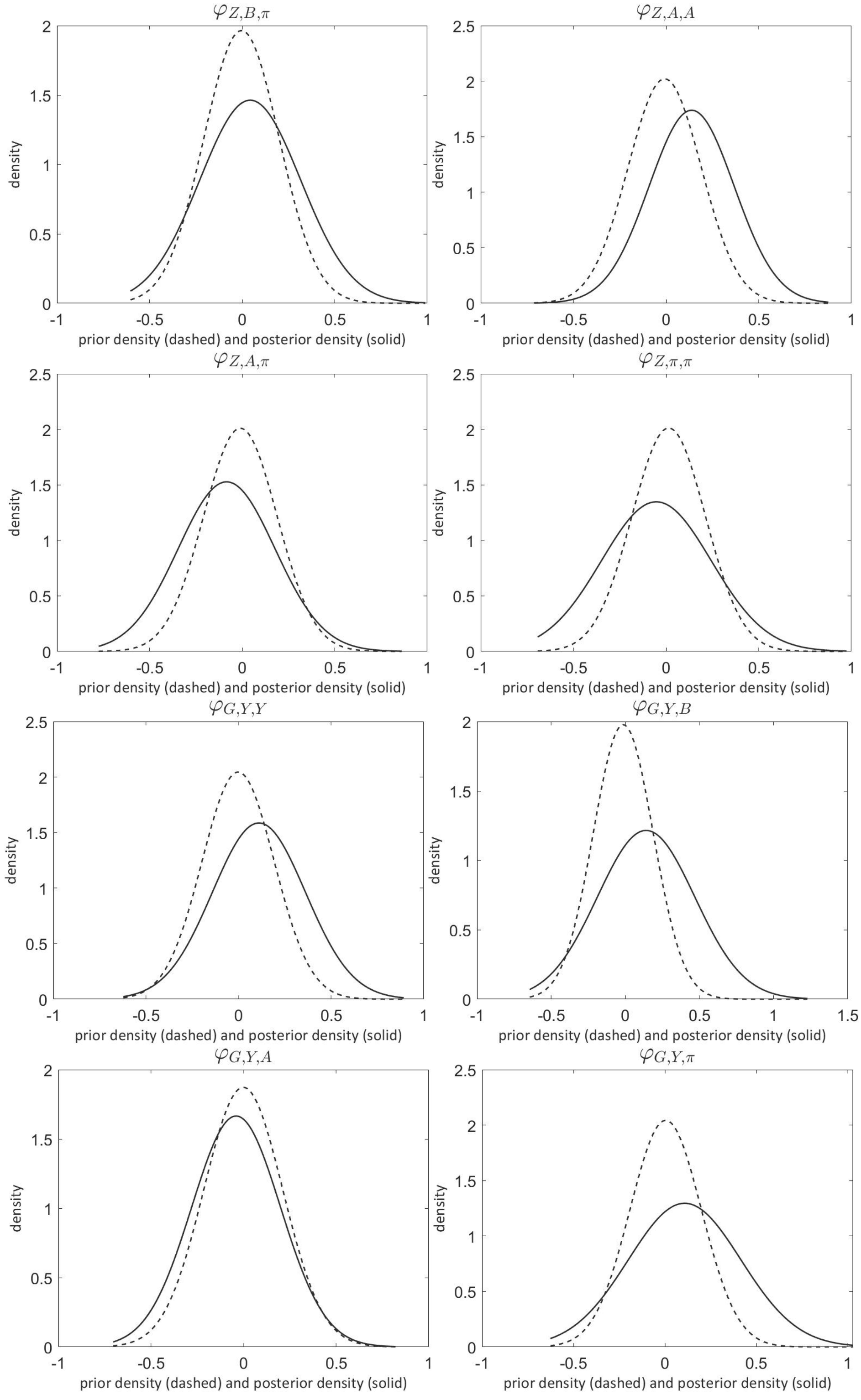
Notes: Posterior and prior density plots for linear fiscal rule parameters. Dashed line corresponds to prior density and solid line to posterior density.

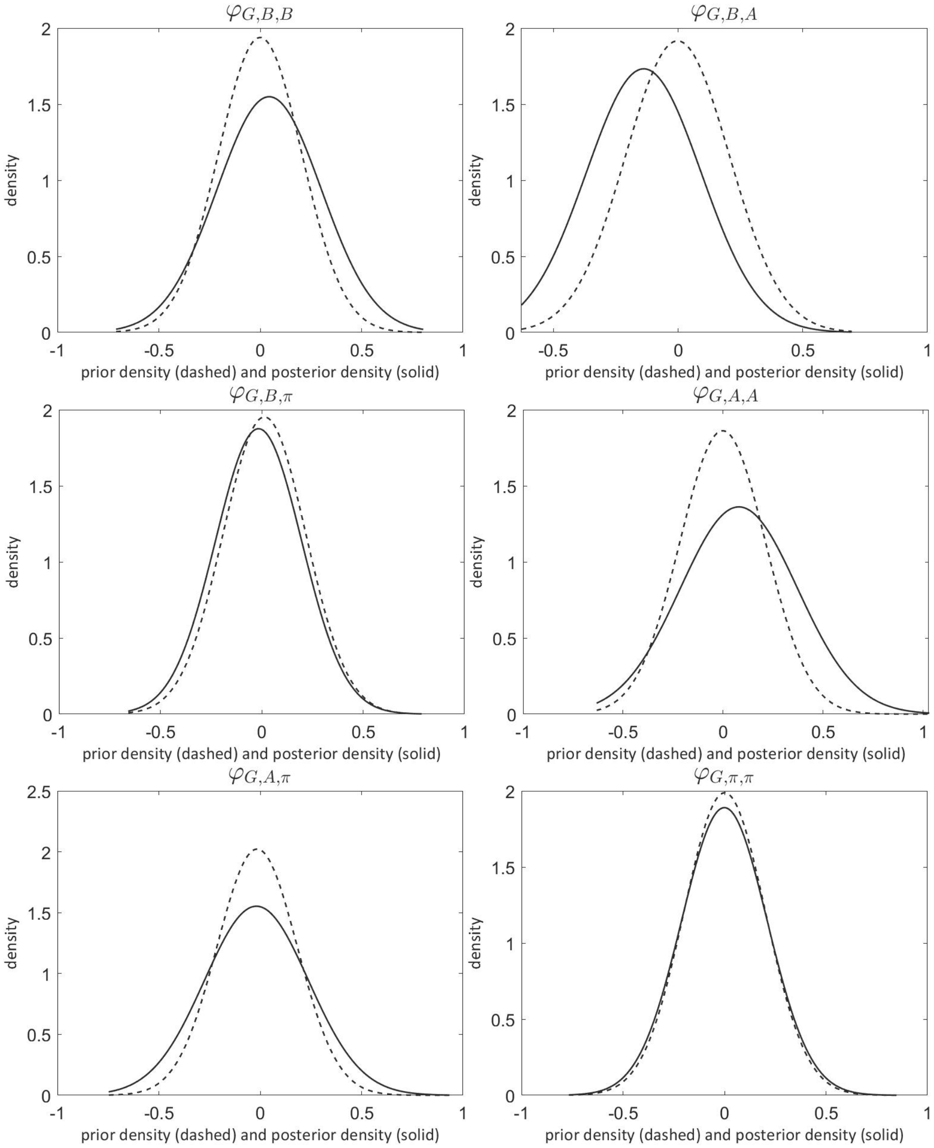
B.3.3 Non-linear rule parameters

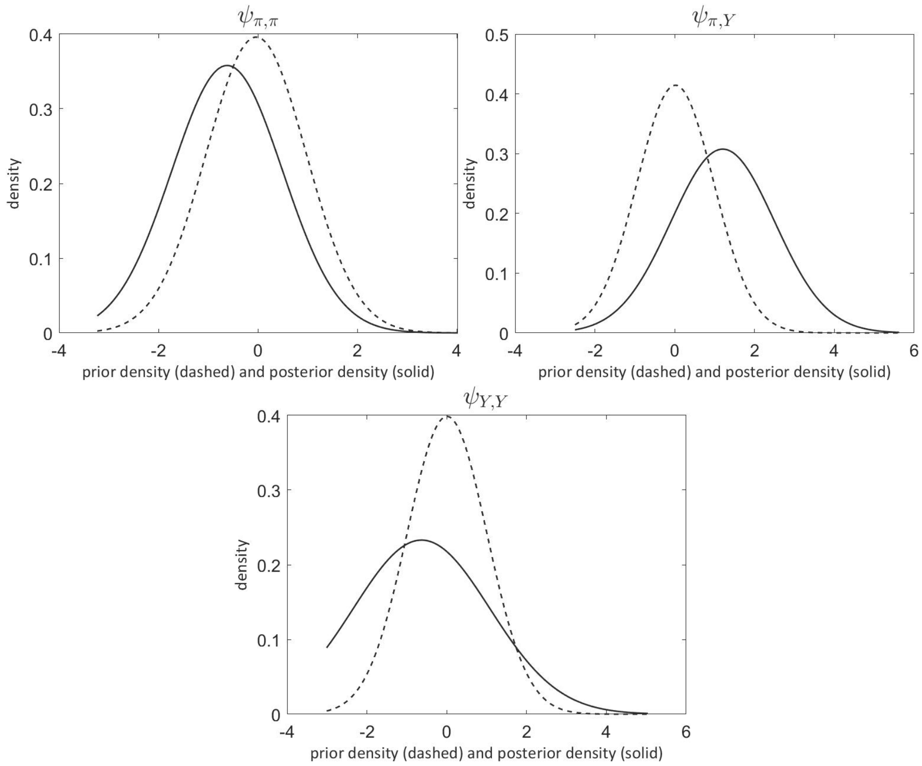
Fig. B.3: Posterior density graphs for non-linear rule parameters




$\varphi_{\tau^l,A,\pi}$
$\varphi_{\tau^l,\pi,\pi}$
$\varphi_{Z,Y,Y}$
$\varphi_{Z,Y,B}$
$\varphi_{Z,Y,A}$
$\varphi_{Z,Y,\pi}$
$\varphi_{Z,B,B}$
$\varphi_{Z,B,A}$
density
prior density (dashed) and posterior density (solid)


$\varphi_{Z,B,\pi}$
$\varphi_{Z,A,A}$
$\varphi_{Z,A,\pi}$
$\varphi_{Z,\pi,\pi}$
$\varphi_{G,Y,Y}$
$\varphi_{G,Y,B}$
$\varphi_{G,Y,A}$
$\varphi_{G,Y,\pi}$
density
prior density (dashed) and posterior density (solid)


$\varphi_{G,B,B}$
$\varphi_{G,B,A}$
$\varphi_{G,B,\pi}$
$\varphi_{G,A,A}$
$\varphi_{G,A,\pi}$
$\varphi_{G,\pi,\pi}$
density
prior density (dashed) and posterior density (solid)



Notes: Posterior and prior density plots for non-linear rule parameters. Dashed line corresponds to prior density and solid line to posterior density.

B.4 Amisano and Tristani (2010) re-estimation

This appendix presents re-estimation results for the Amisano and Tristani (2010) model, which builds the core of the model developed for this chapter and Amisano and Tristani (2010) also designed the conditional particle filter employed here. This section features three estimations. Firstly and secondly, I estimate the Amisano and Tristani (2010) model estimations for the linear and non-linear versions using their base estimation procedure described in their paper based on code developed for this chapter. That includes using the Metropolis-Hastings algorithm on a set of transformed parameters and generating 55,000 draws using a Gaussian transition kernel. The chain is initialized based on a gaussian approximation of the posterior using a preliminary linear run. Further, the first 5000 draws are discarded. The transition kernel is based on the approximated covariance matrix from an initial linear run. Lastly, as a proof of concept, I estimate the non-linear Amisano and Tristani (2010) model using the SMC algorithm described above using the prior distribution of θ, $p(\theta)$, as the initial distribution, $\mu_1(\theta)$. The number of parameter particles is set to 3000, the number of blocks is three and mixture weights and particle degeneracy are as in main estimation. In both applications, the particle filter is initialized based on the unconditional linear distribution to match Amisano and Tristani (2010) who use the linear Covariance matrix. However, I do not use the non-linear unconditional mean and instead use the linear unconditional mean. For this exercise, I rely on the original data set which was available to me as Amisano and Tristani (2010) very kindly shared their code with me.

Table B.2 below represents the estimation results for the posterior simulation for the linear and quadratic approximations on the original dataset. For the linear estimation, all parameters are contained in the highest posterior density intervals produced by Amisano and Tristani (2010). Additionally, all estimates are within 0.4 standard deviations of the original. That is the case even for very tightly measured parameters like β. The highest difference between estimates can be found in the parameters β and ψ_π. For ψ_π the posterior mean is 2.013. In comparison, Amisano and Tristani (2010) estimate the parameter to be slightly lower at 1.947. This difference constitutes to roughly 0.38 standard deviations. Looking at transition plots for the parameter movement during the simulation, parameter transitions appear stable. Furthermore, the estimates appear stable across multiple runs with negligible differences.

Table B.2: Replication Results MH

para	linear		quadratic	
	mean	Sd.	mean	Sd.
β	0.994388	0.00104	0.99311	0.00112
$\gamma-1$	2.35549	0.88414	2.58380	0.82127
h	0.461389	0.06397	0.46263	0.06481
$\phi-1$	4.004861	1.22902	3.44992	0.87042
$\theta-1$	5.484073	2.09163	4.17055	1.42953
ζ	0.402195	0.07384	0.49058	0.07023
l	0.083153	0.04296	0.06354	0.03618
$\psi_\pi-1$	1.01288	0.16845	0.91251	0.15902
ψ_y	0.043456	0.03108	0.06423	0.04667
p_i	0.894975	0.01371	0.89056	0.01343
p_τ	0.506594	0.15723	0.54403	0.14531
p_a	0.997937	0.00167	0.99853	0.00126
p_π	0.988598	0.00711	0.97366	0.01117
σ_τ	0.045471	0.01507	0.04113	0.01340
σ_a	0.013328	0.00162	0.01487	0.00199
σ_π	0.001398	0.00019	0.00133	0.00019
σ_i	0.001941	0.00013	0.00194	0.00013
τ	0.455751	0.28511	0.36558	0.22613
$\pi-1$	0.005609	0.00311	0.00906	0.00318

Notes: Posterior estimates for the linear and non-linear model using the MH algorithm

For the non-linear estimation, the results are very similar. All estimates are contained in the original highest posterior density intervals and within one standard deviation from their posterior estimates rounding up. The larger deviations can be found for the parameters p_i and σ_π. p_i is estimated to be slightly higher at 0.89 in comparison to 0.85. This difference constitutes to approximately 0.8 standard deviations. Similarly, the posterior mean for σ_π is 0.0014 in comparison to the estimate of 0.00174 found by Amisano and Tristani (2010). Overall, the estimates presented here are much closer to Amisano and Tristani (2007) where estimates match within 0.4 standard deviations.

Table B.3 below presents estimation results of the SMC procedure without adaptive tempering on the original Amisano and Tristani (2010) model. The SMC estimation conducted here produces posterior estimates entirely consistent with the original estimates based on the MCMC technique. All estimates are within one standard deviation of the estimates found by Amisano and Tristani (2010).

Table B.3: Replication Results SMC

para	SMC mean	SMC Sd.
β	0.99352	0.00119
$\gamma - 1$	2.29758	1.00288
h	0.46244	0.07441
$\phi - 1$	3.46547	1.21152
$\theta - 1$	4.40969	1.61064
ζ	0.44653	0.08395
l	0.08090	0.04519
$\psi_\pi - 1$	0.91519	0.19334
ψ_y	0.09522	0.06507
p_i	0.88754	0.01787
p_τ	0.49547	0.18285
p_a	0.99823	0.00154
p_π	0.97886	0.01137
σ_τ	0.04094	0.01797
σ_a	0.01421	0.00208
σ_π	0.00126	0.00021
σ_i	0.00197	0.00016
τ	0.38494	0.36830
$\pi - 1$	0.00970	0.00402

Notes: Posterior estimates for the non-linear model using the SMC algorithm

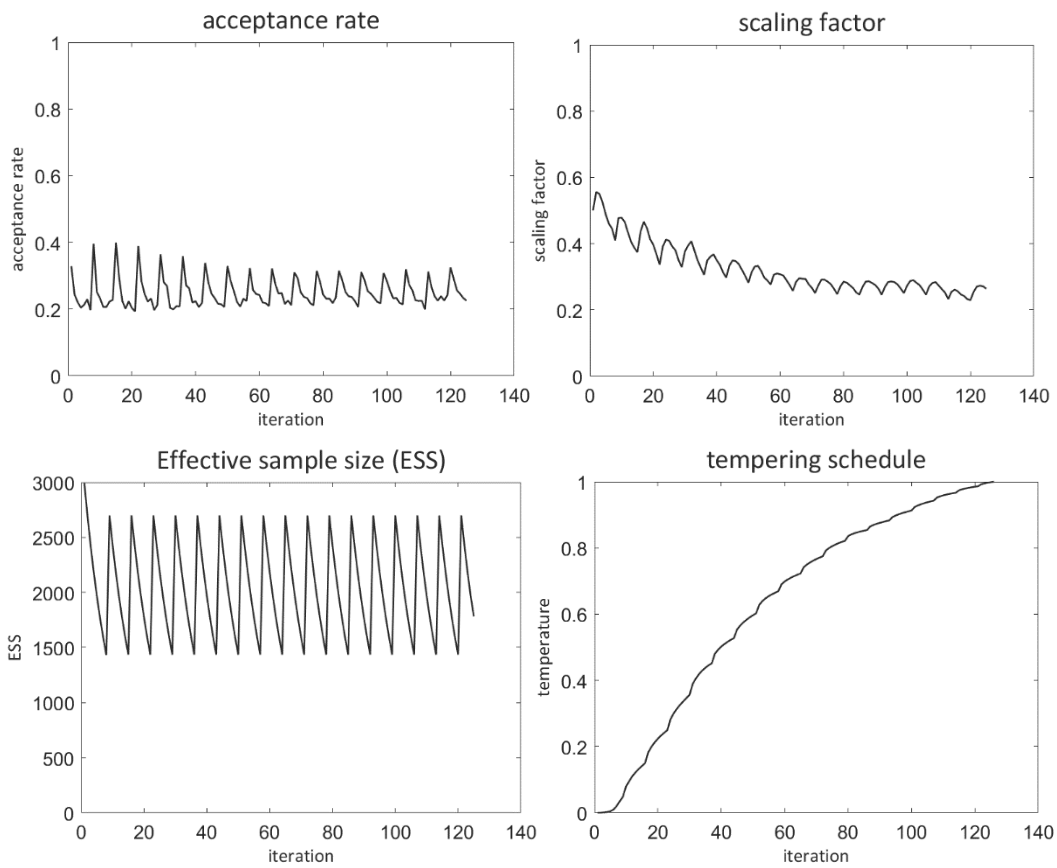
B.5 Estimation diagnostics

Fig. B.4 below displays estimation diagnostics for the main estimation in this chapter. The key mechanic in the diagnostics below comes from the effective sample size dynamic in SMC algorithms. In a SMC algorithm, one generates repeated importance sampling distributions for a sequence of increasingly different distributions. Samples generated for an initial distribution may not match later distributions particularly well and therefore, the number of effective samples will naturally decrease. To control this and to ensure the effective sample size stays high enough, two strategies are employed in this paper. Firstly, Jasra et al (2010) propose a mechanism by which the rate of decay can be controlled. Secondly, once the sample size falls below a threshold, the draws are resampled to create more evenly weighted draws. For more detail on this see the estimation procedure section. The result is the pattern of the effective sample size in fig. 4 below. Fig. 4 shows repeated phases of very consistent decay of the ESS followed by an abrupt upwards jump as a result of the resampling step once the ESS falls below the threshold. The consistency of the behaviour shows how effective the procedure developed by Jasra et al (2010) is at controlling the path of the ESS. As a by-product of the resampling step, one typically receives a much more well behaved distribution of particles in a potentially higher average density area. Therefore, the acceptance rate jump upwards after every resampling step. To ensure that the acceptance rate is close to the target value of 0.24, the Herbst and Schorfheide (2016) target function will then gradually raise the scaling factor for the Metropolis Hastings step.

In the tempering schedule, ϕ_n moves from zero to one. If ϕ_n is equal to zero, then the particle system represents the initial distribution. As ϕ_n moves to one, the SMC samples approximates distributions increasingly more similar to the posterior which culminates at $\phi_n = 1$ with the posterior. The tempering schedules here is concave. This is the case for arguably two reasons. Firstly, the SMC sampler does not start out in an area of low density as the initial distribution is an approximated posterior. Therefore, the sampler can add information quickly without generating bridge distributions which are too dissimilar. Secondly, as the temperature increases, the noise of the target density due to the particle filter approximation increases.[24] With increasing noise in the target function, the current posterior become more difficult to transverse and the speed of the tempering schedule decreases as it has to ensure the planned decay of the ESS.

[24] Assume $p(y|\theta)$ is distributed as a normal with mean, $\mu_{p(y|\theta)}$, and variance, $\sigma^2_{p(y|\theta)}$. Then the variance of the tempered distribution is defined as: $Var(\pi_n(\theta)) \propto Var\left(\left(p(y|\theta)p(\theta)\right)^{\phi_n}\mu_1(\theta)^{1-\phi_n}\right) = \left(\mu_1(\theta)^{1-\phi_n}\left(p(\theta)\right)^{\phi_n}\right)^2 VAR\left(\left(p(y|\theta)\right)^{\phi_n}\right)$. For small ϕ_n,$VAR\left(\left(p(y|\theta)\right)^{\phi_n}\right)$ is essentially zero. As ϕ_n increases, $VAR\left(\left(p(y|\theta)\right)^{\phi_n}\right)$ increases reaching the full variance at $\phi_n = 1$

Fig. B.4: Simulation Diagnostics



Notes: Simulation Diagnostics for the mean estimation. It includes acceptance rates, scaling factor, effective samples size and temperature path across the iteration.

B.6 Regression Tables for IRFs

Table B.4: Regression of IRFs on impact of output to government consumption and transfer shocks on initial conditions

variable	$IRF_{t+4}^{Y}\|v^{G}$			$IRF_{t+4}^{Y}\|v^{Z}$		
	estim.	std.	t-val.	estim.	std.	t-val.
$\tilde{\pi}_{t-1}$	-0.0050	0.0003	-16.92	-0.0093	0.0007	-12.61
$\tilde{Y}_{t-1}$	-0.0003	8.52E-06	-30.24	0.0019	2.16E-05	89.09
$\tilde{\imath}_{t-1}$	-0.0061	0.0002	-25.29	-0.0128	0.0006	-21.10
$\tilde{B}_{t-1}$	-7.27E-05	7.81E-07	-93.17	-0.0002	1.95E-06	-79.19
$\tilde{\tau}_{t-1}^{l}$	0.0003	5.57E-06	46.07	0.0004	1.39E-05	31.88
$\tilde{Z}_{t-1}$	-7.88E-05	4.47E-06	-17.62	-0.0002	1.11E-05	-20.20
$\tilde{G}_{t-1}$	0.0010	3.66E-06	260.76	9.58E-05	9.15E-06	10.46
$\tilde{a}_{t-1}$	-0.0003	2.65E-06	-94.69	-0.0007	6.74E-06	-110.01
$\tilde{\tau}_{t-1}^{c}$	3.47E-05	4.92E-06	7.06	8.06E-05	1.28E-05	6.27
π_{t-1}^{*}	0.0043	0.0006	7.47	0.0042	0.0014	2.91
Const.	0.0957	0.0001	708.11	-0.0085	0.0003	-25.16
R^2	0.8789			0.6489		
$RMSE_{lin}$	0.0331			0.0830		
$RMSE_{mean}$	0.0951			0.1400		
obs.	60000			60000		

Notes: Regressions of IRF_{t+1}^{Y} on initial conditions for a positive, one standard deviation shock to government consumption and transfers, respectively. Initial conditions are phrased as percentage steady state deviations as per the model set-up. $RMSE_{lin}$ is the in-sample root mean square error of the full linear model and $RMSE_{mean}$ is the RMSE for a mean model.

Table B.5: Regression of IRFs on impact of output to consumption and labour tax shocks on initial conditions

variable	$IRF^{Y}_{t+4}\|v^{\tau^c}$			$IRF^{Y}_{t+4}\|v^{\tau^l}$		
	estim.	std.	t-val.	estim.	std.	t-val.
$\tilde{\pi}_{t-1}$	-0.0020	9.30E-05	-21.194	-0.0054	0.0003	-18.12
$\tilde{Y}_{t-1}$	0.0002	2.69E-06	81.743	0.0006	8.67E-06	66.11
$\tilde{\iota}_{t-1}$	-0.0015	7.61E-05	-19.770	-0.0052	0.0002	-21.00
$\tilde{B}_{t-1}$	-2.56E-05	2.46E-07	-103.990	-8.33E-05	7.91E-07	-105.30
$\tilde{\tau}^{l}_{t-1}$	5.76E-05	1.75E-06	32.842	0.0005	5.70E-06	82.46
$\tilde{Z}_{t-1}$	-1.02E-05	1.40E-06	-7.279	-4.36E-05	4.54E-06	-9.60
$\tilde{G}_{t-1}$	-5.51E-06	1.15E-06	-4.796	6.92E-06	3.70E-06	1.87
$\tilde{a}_{t-1}$	-8.29E-05	8.47E-07	-97.844	-0.0002	2.69E-06	-75.74
$\tilde{\tau}^{c}_{t-1}$	9.04E-05	1.58E-06	57.057	2.76E-05	5.21E-06	5.30
π^{*}_{t-1}	0.0024	0.0002	13.098	0.0071	0.0006	12.09
Const.	0.0065	4.26E-05	151.902	0.0353	0.0001	256.67
R^2	0.6929			0.5345		
$RMSE_{lin}$	0.0104			0.0337		
$RMSE_{mean}$	0.0188			0.0494		
obs.	60000			60000		

Notes: Regressions of IRF^{Y}_{t+1} on initial conditions for a positive, one standard deviation shock to consumption and labour taxation, respectively. Initial conditions are phrased as percentage steady state deviations as per the model set-up. $RMSE_{lin}$ is the in-sample root mean square error of the full linear model and $RMSE_{mean}$ is the RMSE for a mean model.

Appendix C

Appendix to Chapter 3

C.1 Additional forecasting performance result tables

Table C.1: Forecasting breakdown for output by samples for the rolling window

		Sample: 1962-1999		Sample: 2000-end	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	t + 1	0.010	0.000	0.016	0.000
BVAR(1) Min	t + 1	0.010	-2.832	0.016	-0.564
BVAR(1) $\lambda = 1$	t + 1	0.009	-7.851	0.015	-7.842
BVAR(1) $\lambda = 2$	t + 1	0.009	-4.961	0.015	-8.330
VAR(1)	t + 4	0.028	0.000	0.026	0.000
BVAR(1) Min	t + 4	0.027	-2.655	0.026	-1.388
BVAR(1) $\lambda = 1$	t + 4	0.026	-5.367	0.027	2.829
BVAR(1) $\lambda = 2$	t + 4	0.026	-7.186	0.026	-2.348
VAR(4)	t + 1	0.020	100.866	0.038	138.745
BVAR(4) Min	t + 1	0.011	6.718	0.021	33.101
BVAR(4) $\lambda = 1$	t + 1	0.011	10.412	0.028	73.104
BVAR(4) $\lambda = 2$	t + 1	0.010	-1.053	0.017	6.891
VAR(4)	t + 4	0.060	118.419	0.352	1232.541
BVAR(4) Min	t + 4	0.028	2.938	0.029	8.040
BVAR(4) $\lambda = 1$	t + 4	0.037	34.759	0.133	401.158
BVAR(4) $\lambda = 2$	t + 4	0.032	17.011	0.030	15.063
ZLB	t + 1	0.009	-8.938	0.017	4.334
ZLB	t + 4	0.026	-7.100	0.028	4.113

Notes: Breakdown of the forecasting performance of output for the DSGE model for the rolling window estimation broken down into two subsamples: 1962 to 1999 and 2000 to 2021. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the two subsamples presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Table C.2: Forecasting breakdown for output by samples for the expanding window

		Sample: 1962-1999		Sample: 2000-end	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.009	0.000	0.015	0.000
BVAR(1) Min	$t+1$	0.009	-1.758	0.015	0.031
BVAR(1) $\lambda = 1$	$t+1$	0.009	-5.192	0.014	-5.447
BVAR(1) $\lambda = 2$	$t+1$	0.009	-3.326	0.014	-5.945
VAR(1)	$t+4$	0.024	0.000	0.038	0.000
BVAR(1) Min	$t+4$	0.024	-0.121	0.038	-0.021
BVAR(1) $\lambda = 1$	$t+4$	0.020	-14.917	0.029	-23.954
BVAR(1) $\lambda = 2$	$t+4$	0.021	-10.869	0.026	-30.727
VAR(4)	$t+1$	0.011	24.531	0.018	16.175
BVAR(4) Min	$t+1$	0.010	6.085	0.017	8.507
BVAR(4) $\lambda = 1$	$t+1$	0.009	0.041	0.017	8.573
BVAR(4) $\lambda = 2$	$t+1$	0.009	-0.987	0.016	7.513
VAR(4)	$t+4$	0.038	57.754	0.039	3.911
BVAR(4) Min	$t+4$	0.028	17.655	0.039	3.719
BVAR(4) $\lambda = 1$	$t+4$	0.025	4.264	0.033	-13.413
BVAR(4) $\lambda = 2$	$t+4$	0.025	3.953	0.030	-20.374
ZLB	$t+1$	0.009	-5.175	0.017	8.739
ZLB	$t+4$	0.025	2.848	0.026	-30.216

Notes: Breakdown of the forecasting performance of output for the DSGE model for the expanding window estimation broken down into two subsamples: 1962 to 1999 and 2000 to 2021. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the two subsamples presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Table C.3: Forecasting breakdown for debt by samples for the rolling window

		Sample: 1962-1999		Sample: 2000-end	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.036	0.00	0.031	0.00
BVAR(1) Min	$t+1$	0.035	-0.90	0.031	-0.09
BVAR(1) $\lambda = 1$	$t+1$	0.034	-5.80	0.031	-1.22
BVAR(1) $\lambda = 2$	$t+1$	0.033	-5.99	0.031	-0.17
VAR(1)	$t+4$	0.078	0.00	0.063	0.00
BVAR(1) Min	$t+4$	0.078	-1.21	0.063	0.05
BVAR(1) $\lambda = 1$	$t+4$	0.065	-17.60	0.063	0.75
BVAR(1) $\lambda = 2$	$t+4$	0.064	-17.95	0.062	-0.51
VAR(4)	$t+1$	0.042	18.26	0.033	3.57
BVAR(4) Min	$t+1$	0.035	-1.98	0.031	-1.92
BVAR(4) $\lambda = 1$	$t+1$	0.033	-6.91	0.031	-2.01
BVAR(4) $\lambda = 2$	$t+1$	0.033	-7.98	0.030	-3.96
VAR(4)	$t+4$	0.092	17.50	0.059	-6.03
BVAR(4) Min	$t+4$	0.074	-6.00	0.062	-1.79
BVAR(4) $\lambda = 1$	$t+4$	0.067	-14.91	0.056	-11.50
BVAR(4) $\lambda = 2$	$t+4$	0.064	-18.19	0.057	-9.68
ZLB	$t+1$	0.032	-9.28	0.031	-1.46
ZLB	$t+4$	0.061	-21.64	0.062	-1.60

Notes: Breakdown of the forecasting performance of debt for the DSGE model for the rolling window estimation broken down into two subsamples: 1962 to 1999 and 2000 to 2021. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the two subsamples presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Table C.4: Forecasting breakdown for gov. consumption by samples for the rolling window

		Sample: 1962-1999		Sample: 2000-end	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.040	0.00	0.047	0.00
BVAR(1) Min	$t+1$	0.040	0.32	0.046	-1.17
BVAR(1) $\lambda=1$	$t+1$	0.037	-6.49	0.037	-22.00
BVAR(1) $\lambda=2$	$t+1$	0.038	-4.87	0.036	-22.94
VAR(1)	$t+4$	0.094	0.00	0.105	0.00
BVAR(1) Min	$t+4$	0.090	-4.41	0.106	0.72
BVAR(1) $\lambda=1$	$t+4$	0.085	-9.99	0.098	-6.58
BVAR(1) $\lambda=2$	$t+4$	0.085	-9.55	0.095	-9.45
VAR(4)	$t+1$	0.076	92.05	0.080	69.92
BVAR(4) Min	$t+1$	0.042	4.75	0.044	-5.44
BVAR(4) $\lambda=1$	$t+1$	0.042	5.46	0.049	4.22
BVAR(4) $\lambda=2$	$t+1$	0.038	-3.57	0.042	-9.77
VAR(4)	$t+4$	0.141	49.24	0.156	48.17
BVAR(4) Min	$t+4$	0.091	-3.67	0.120	13.70
BVAR(4) $\lambda=1$	$t+4$	0.087	-7.56	0.115	9.52
BVAR(4) $\lambda=2$	$t+4$	0.083	-11.75	0.104	-1.23
ZLB	$t+1$	0.035	-12.24	0.035	-25.62
ZLB	$t+4$	0.077	-18.81	0.085	-18.94

Notes: Breakdown of the forecasting performance of government consumption for the DSGE model for the rolling window estimation broken down into two subsamples: 1962 to 1999 and 2000 to 2021. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the two subsamples presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Table C.5: Forecasting breakdown for debt by samples for the expanding window

model	step	Sample: 1962-1999 RMSE	Sample: 1962-1999 pct. dev. to VAR(1)	Sample: 2000-end RMSE	Sample: 2000-end pct. dev. to VAR(1)
VAR(1)	$t+1$	0.021	0.00	0.021	0.00
BVAR(1) Min	$t+1$	0.021	-0.33	0.021	0.84
BVAR(1) $\lambda = 1$	$t+1$	0.022	4.83	0.020	-4.93
BVAR(1) $\lambda = 2$	$t+1$	0.022	3.90	0.020	-3.12
VAR(1)	$t+4$	0.069	0.00	0.074	0.00
BVAR(1) Min	$t+4$	0.068	-0.94	0.074	0.94
BVAR(1) $\lambda = 1$	$t+4$	0.064	-6.72	0.075	1.73
BVAR(1) $\lambda = 2$	$t+4$	0.060	-12.04	0.079	7.03
VAR(4)	$t+1$	0.036	70.70	0.050	140.83
BVAR(4) Min	$t+1$	0.021	2.16	0.023	11.58
BVAR(4) $\lambda = 1$	$t+1$	0.024	14.68	0.024	16.34
BVAR(4) $\lambda = 2$	$t+1$	0.020	-4.56	0.020	-3.80
VAR(4)	$t+4$	0.120	74.58	0.500	579.41
BVAR(4) Min	$t+4$	0.067	-2.13	0.092	24.71
BVAR(4) $\lambda = 1$	$t+4$	0.074	7.55	0.097	31.20
BVAR(4) $\lambda = 2$	$t+4$	0.062	-10.11	0.084	14.16
ZLB	$t+1$	0.019	-8.69	0.021	0.09
ZLB	$t+4$	0.054	-21.29	0.066	-9.66

Notes: Breakdown of the forecasting performance of debt for the DSGE model for the expanding window estimation broken down into two subsamples: 1962 to 1999 and 2000 to 2021. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the two subsamples presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

Table C.6: Forecasting breakdown for gov. consumption by samples for the expanding window

		Sample: 1962-1999		Sample: 2000-end	
model	step	RMSE	pct. dev. to VAR(1)	RMSE	pct. dev. to VAR(1)
VAR(1)	$t+1$	0.009	0.00	0.015	0.00
BVAR(1) Min	$t+1$	0.009	-1.76	0.015	0.03
BVAR(1) $\lambda=1$	$t+1$	0.009	-5.19	0.014	-5.45
BVAR(1) $\lambda=2$	$t+1$	0.009	-3.33	0.014	-5.95
VAR(1)	$t+4$	0.024	0.00	0.038	0.00
BVAR(1) Min	$t+4$	0.024	-0.12	0.038	-0.02
BVAR(1) $\lambda=1$	$t+4$	0.020	-14.92	0.029	-23.95
BVAR(1) $\lambda=2$	$t+4$	0.021	-10.87	0.026	-30.73
VAR(4)	$t+1$	0.011	24.53	0.018	16.18
BVAR(4) Min	$t+1$	0.010	6.09	0.017	8.51
BVAR(4) $\lambda=1$	$t+1$	0.009	0.04	0.017	8.57
BVAR(4) $\lambda=2$	$t+1$	0.009	-0.99	0.016	7.51
VAR(4)	$t+4$	0.038	57.75	0.039	3.91
BVAR(4) Min	$t+4$	0.028	17.66	0.039	3.72
BVAR(4) $\lambda=1$	$t+4$	0.025	4.26	0.033	-13.41
BVAR(4) $\lambda=2$	$t+4$	0.025	3.95	0.030	-20.37
ZLB	$t+1$	0.009	-5.18	0.017	8.74
ZLB	$t+4$	0.025	2.85	0.026	-30.22

Notes: Breakdown of the forecasting performance of government consumption for the DSGE model for the expanding window estimation broken down into two subsamples: 1962 to 1999 and 2000 to 2021. The first column includes model descriptions for the set of models considered. The second column defines the forecast horizon. This a followed by two columns each for the two subsamples presenting estimates for the Root mean square error (RMSE) and percentage deviations to the VAR(1).

C.2 Occbin solution strategy for Klein (2000) system set up

To implement the Zero Lower Bound (ZLB) estimation for this chapter, I rely on a variation of the Occbin approach developed in Guerrieri and Iacoviello (2015) adapted to the Klein (2000) setup. The DSGE literature has independently developed several solvers for DSGE models like the Klein (2000), Sims (2002), or the solver developed for Dynare, among others. Each representation has its own individual advantages for certain models, data structures, computational speed or accuracy. However, most of them come with their own system setup and nuanced differences in the way the solution strategies play out. A lot of the system setup is closely related, and systems can be reorganized (for a brilliant survey on DSGE solver, see Anderson (2007), which also offers detail on the system relations of various approaches). The problem is that it typically leaves you to expand the system size quite drastically, which in turn increases the runtime of simulations.

Further, a lot of methods that build on the basic representations and solvers are adapted around the original notation preferred by the respective economists based on the problem at hand, as is the case for Occbin. In particular, it is the case that the original notation offers a favourable solution representation. For this chapter, the Klein (2000) approach is favourited for several reasons. Firstly, the Klein (2002) is a robust and fast DSGE solver. Secondly, the model files were written for chapter 2 for the Klein (2000) set up and re-writing DSGE files is an incredibly error-prone procedure. Lastly, while Dynare delivers a fantastic ecosystem for writing, developing and estimating DSGE models, its codebase is much more difficult to adapt to other problems where you don't fully make use of the Dynare codebase. Therefore, in this section, I describe how to adjust the Klein (2000) setup to solve for the ZLB recursive model structure as in Guerrieri and Iacoviello (2015). Crucially, Klein (2000) distinguishes between predetermined and non-predetermined variables, while Guerrieri and Iacoviello (2015) does not.

Guerrieri and Iacoviello (2015) begin by noting that the model with occasionally binding constraints can be in one of two phases. Either the occasionally binding constraints are slack, or they are binding today. If the constraint is slack, then the economy is defined to be in the reference regime and governed by a set of equations: $g(X_{t+1}, X_t, X_{t-1}) \leq 0$. Otherwise, the constraint is binding in the alternative regime, and the economy faces an alternative system: $h(X_{t+1}, X_t, X_{t-1}) > 0$. In the linear form, the reference regime and alternative regimes, respectively, can be summarized as follows:

$$AE_tX_{t+1} + BX_t + CX_{t-1} + Dz_{t+1} + Ez_t = 0,$$

$$\tilde{A}E_tX_{t+1} + \widetilde{B}X_t + \tilde{C}X_{t-1} + \widetilde{D}z_{t+1} + \tilde{E}z_t + \tilde{F} = 0,$$

where X_t is of dimension $(n_x \times 1)$ and $z_t$is $(n_z \times 1)$. The matrices A, B and C are sized $(n_x \times n_x)$ whereas D and E are of dimension $(n_z \times n_z)$. Further, $\tilde{F}$ is a $(n_x \times 1)$ column vector.

z_{t+1} and z_t may be reduced to $\boldsymbol{z_t} = [{z_t}', {z_{t+1}}']$' using $\Psi = [E, D]$ to match Guerrieri and Iacoviello (2015) based on Anderson (2007):

$$AE_tX_{t+1} + BX_t + CX_{t-1} + \Psi \boldsymbol{z}_t = 0,$$

$$\tilde{A}E_tX_{t+1} + \widetilde{B}X_t + \tilde{C}X_{t-1} + \widetilde{D}z_{t+1} + \widetilde{\Psi}\boldsymbol{z}_t + \tilde{F} = 0.$$

Further, z_{t+1} typically follows the following type of process:

$$z_{t+1} = \phi z_t + \varepsilon_{t+1}.$$

To move from the Guerrieri and Iacoviello (2015) setup to the Klein (2000) notation, the system has to be rewritten such that the model includes one lead endogenous vector but no lags. To do so, I introduce auxiliary vectors k_t and u_t. k_t is a vector of state variables which includes predetermined variables that may be endogenous at other lags and purely exogenous variables. At time t, X_{t-1} has already been determined and therefore satisfies the predetermination criterion. Therefore, k_t may be constructed as $k_t = [X'_{t-1}, z'_t]'$. k_t is a vector of size $(n_k \times 1)$ where $n_k = n_x + n_z$. u_t is a vector of non-state variables or variables that are endogenous today. u_t is set to X_t. Consequently, the system can be rewritten under the reference and alternative regime, respectively:

$$\begin{bmatrix} B_1 & D & A \\ 0 & I_{n_z} & 0 \\ I_{n_x} & 0 & 0 \end{bmatrix} \begin{bmatrix} X_t \\ z_{t+1} \\ E_tX_{t+1} \end{bmatrix} + \begin{bmatrix} C & E & B_2 \\ 0 & -\phi & 0 \\ 0 & 0 & -I_{n_x} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ z_t \\ X_t \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_{t+1} \\ 0 \end{bmatrix} = 0,$$

$$\begin{bmatrix} \widetilde{B}_1 & \widetilde{D} & \tilde{A} \\ 0 & I_{n_z} & 0 \\ I_{n_x} & 0 & 0 \end{bmatrix} \begin{bmatrix} X_t \\ z_{t+1} \\ E_tX_{t+1} \end{bmatrix} + \begin{bmatrix} \tilde{C} & \tilde{E} & \widetilde{B}_2 \\ 0 & -\phi & 0 \\ 0 & 0 & -I_{n_x} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ \widetilde{z_t} \\ X_t \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{\varepsilon}_{t+1} \\ 0 \end{bmatrix} + \begin{bmatrix} \tilde{F} \\ 0 \\ 0 \end{bmatrix} = 0,$$

In practice, one may wish to reduce the size of k_t by omitting elements of X_{t-1} that do not actually enter the system of equations. In this case, that is not possible, as one of the requirements for the Occbin solution to bind is that there exists a linear mapping from the current predetermined vector to the previous control vector and vice versa. The easiest way to ensure that that is the case is by following the above structure and including redundant predetermined variables. This can be done by simply adding identity equations.

If the economy is in the reference regime, the Klein (2002) algorithm allows us to construct a solution of the following type for the predetermined and exogenous variables:

$$\begin{bmatrix} X_t \\ z_{t+1} \end{bmatrix} = H \begin{bmatrix} X_{t-1} \\ z_t \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_{t+1} \end{bmatrix}, where\ H = \begin{bmatrix} H_{11} & H_{12} \\ 0 & \phi \end{bmatrix}.$$

Alternatively, if the constraint binds, then one can recursively solve for the system today based on a proposed duration of the ZLB, after which the economy remains in the unconstrained case. To illustrate, assume that in $t = 3$, the system is unconstrained, but in $t = 2$ and $t = 1$ the economy is constrained. In $t = 3$, we know that the system transitions by the unconstrained law of motioned defined above and can be expressed with indices as follows:

$$\begin{bmatrix} X_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_4 \end{bmatrix}.$$

In $t = 2$, the constrained system governs the economy:

$$\begin{bmatrix} \widetilde{B} & \widetilde{D} & \tilde{A} \\ 0 & I_{n_z} & 0 \\ I_{n_x} & 0 & 0 \end{bmatrix} \begin{bmatrix} X_2 \\ z_3 \\ E_2 X_3 \end{bmatrix} + \begin{bmatrix} \tilde{C} & \tilde{E} & 0 \\ 0 & -\phi & 0 \\ 0 & 0 & -I_{n_x} \end{bmatrix} \begin{bmatrix} X_1 \\ z_2 \\ X_2 \end{bmatrix} + \begin{bmatrix} \tilde{F} \\ 0 \\ 0 \end{bmatrix}.$$

To solve this system, $E_2 X_3$ needs to be evaluated. The agent knows that in $t = 3$, the economy is in the reference regime, and because we have defined the law of motion in that case above, we can evaluate $E_2 X_3$ as:

$$E_2 X_3 = E_2(H_{11} X_2 + H_{12} z_3) = (H_{11} X_2 + H_{12} \phi z_2).$$

Then, we first take the conditional expectation, E_2, of the entire system, and we can substitute $E_2 X_3$ out and solve for X_2 as a function of X_1 and z_2:

$$\widetilde{B} X_2 + \widetilde{D} \phi z_2 + \tilde{A}(H_{11} X_2 + H_{12} \phi z_2) + \tilde{C} X_1 + \tilde{E} z_2 + \tilde{F} = 0,$$

$$X_2 = -\left(\widetilde{B} + \tilde{A} H_{11}\right)^{-1} \tilde{C} X_1 - \left(\widetilde{B} + \tilde{A} H_{11}\right)^{-1} \left(\widetilde{D}\phi + \tilde{A} H_{12} \phi + \tilde{E}\right) z_2 - \left(\widetilde{B} + \tilde{A} H_{11}\right)^{-1} \tilde{F}.$$

At this point, the equation above defines a comparable recursive solution to Guerrieri and Iacoviello (2015). The only difference is the explicit inclusion of z_2. Vectorizing allows us to define the following system based on the Klein (2000) setup:

$$\begin{bmatrix} X_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} -\left(\widetilde{B} + \tilde{A} H_{11}\right)^{-1} \tilde{C} & -\left(\widetilde{B} + \tilde{A} H_{11}\right)^{-1} (\widetilde{D}\phi + \tilde{A} H_{12} \phi + \tilde{E}) \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} -\left(\widetilde{B} + \tilde{A} H_{11}\right)^{-1} \tilde{F} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_3 \end{bmatrix},$$

$$\begin{bmatrix} X_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} H_{11,t=2} & H_{12,t=2} \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} R_{1,t=2} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_3 \end{bmatrix}.$$

Similarly, in $t = 1$ we can substitute again using $H_{11,t=2}$, $H_{12,t=2}$ and $R_{1,t=2}$ to recover:

$$\begin{bmatrix} X_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} -\left(\widetilde{B} + \tilde{A}H_{11,t=2}\right)^{-1}\widetilde{C} & -\left(\widetilde{B} + \tilde{A}H_{11,t=2}\right)^{-1}(\widetilde{D}\phi + \tilde{A}H_{12,t=2}\phi + \tilde{E}) \\ 0 & \phi \end{bmatrix} \begin{bmatrix} X_0 \\ z_1 \end{bmatrix}$$
$$+ \begin{bmatrix} -\left(\widetilde{B} + \tilde{A}H_{11,t=2}\right)^{-1}(\tilde{A}R_{1,t=2} + \tilde{F}) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_2 \end{bmatrix}$$

C.3 Average posterior estimates of the DSGE-VAR model

In this section, I present estimates of the structural DSGE parameters constructed as a by-product of the $DSGE - VAR(1)$ rolling window estimations with λ equal to two. To construct a more direct comparison between the estimations in chapter two and three that is less sample dependent, I construct mean estimates and the associated standard deviations across all rolling window estimations. The results are presented in the tables below.

Table C.7: Averaged rolling window core economic parameter estimates

para	mean	sd.	para	mean	sd.
β	0.99565	0.00320	p_π	0.90433	0.06077
$\gamma - 1$	5.08352	2.86637	σ_{τ^l}	0.00902	0.00494
h	0.46517	0.21144	σ_{τ^c}	0.01341	0.00608
$\phi - 1$	0.99182	0.59086	σ_T	0.03331	0.01204
$\theta - 1$	2.02857	2.01873	σ_G	0.00936	0.00281
ζ	0.69881	0.11944	σ_a	0.01693	0.00714
l	0.25996	0.14280	σ_i	0.00076	0.00038
$\psi_\pi - 1$	0.92541	0.20987	σ_π	0.00053	0.00034
ψ_y	0.18667	0.08098	τ^l	0.22996	0.00108
p_{τ^l}	0.59955	0.21354	τ^c	0.01576	0.00160
p_{τ^c}	0.72088	0.14795	s_g	0.05974	0.00108
p_Z	0.81896	0.22103	s_b	0.51739	0.02349
p_G	0.62932	0.21554	π	0.00560	0.00027
p_a	0.99675	0.04054			
p_i	0.81442	0.08694			

Table C.8: Averaged rolling window fiscal parameter estimates

para	mean	sd.
$\mu_{\tau^l,Y}$	0.25487	0.23454
$\mu_{\tau^l,B}$	0.37331	0.45616
$\mu_{\tau^l,A}$	0.02876	0.15617
$\mu_{\tau^l,\pi}$	0.01102	0.11085
$\mu_{Z,Y}$	0.15466	0.10821
$\mu_{Z,B}$	0.42681	0.58104
$\mu_{Z,A}$	0.02476	0.11608
$\mu_{Z,\pi}$	-0.00583	0.10794
$\mu_{G,Y}$	0.19487	0.16964
$\mu_{G,B}$	0.16836	0.23302
$\mu_{G,A}$	-0.04079	0.14352
$\mu_{G,\pi}$	-0.00208	0.10471

Firstly, parameter uncertainty is almost uniformly higher in chapter three than it is in chapter two. This is a fairly natural by-product of the smaller estimation samples, less posterior draws and the change in estimation technique. Secondly, while a large share of parameters is estimated very similar to chapter 2, some core economic and fiscal policy parameters are estimated quite differently. For example, the parameter of relative risk aversion, $\gamma - 1$, is estimated with a mean of 2.89 in chapter 2, while the DSGE-VAR estimate sits closer to 5. Furthermore, the goods elasticity of substitution, $\theta - 1$, is estimated to be substantially lower at around 2. Similarly, the DSGE-VAR estimates for the inflation indexation are lower at 0.26. Overall, the autoregressive parameters tend to be quite a bit lower. Changes persist also for the fiscal policy parameters where $\mu_{\tau^l,Y}$, $\mu_{\tau^l,B}$ and $\mu_{Z,B}$ are all estimated with higher means in chapter three. Arguably, the difference in estimates mostly arise due to moving from a full likelihood to a quasi-likelihood approach, differences in sample selection and switching away from the non-linear DSGE. The differences in parameter estimates may also explain the differences in impulse response results between chapters two and three.