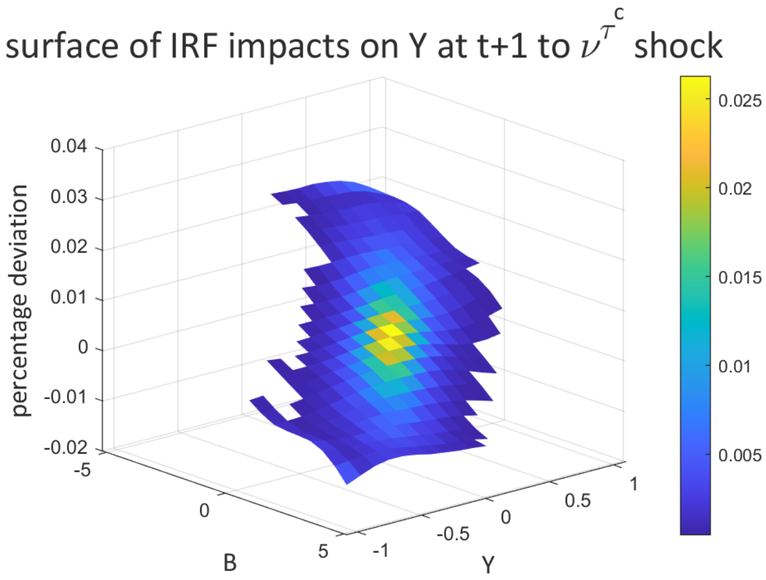
Fiscal policy and the business cycle: An argument for non-linear policy rules

Fleischhacker, Jan

23 October 2024

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



Jan Fleischhacker

Independent researcher

Germany

23/10/2024

Abstract

In this paper, 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 paper 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, 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. This paper forms the second chapter of my PhD thesis completed in 2024 at Royal Holloway, University of London, which can be found here: https://pure.royalholloway.ac.uk/en/publications/essays-on-fiscal-rule-design-and-its-implications

Contents

List of Figures

List of Tables

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

1.1 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 paper 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. 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 paper 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 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 paper 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 paper is structured as follows. Section 1.2 presents a literature review. Section 1.3 sets up the model and presents the dynamic equations. Section 1.4 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 1.5 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.

1.2 Literature review

1.2.1 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.[1] 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

[1] 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).

1.2.2 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.[2] 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

[2] 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:

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

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.

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.

1.3 Model description

The following section describes the New Keynesian model developed in this paper. 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 1.3.1 and 1.3.2 describe the federal government and how the federal government rules change over the business cycle. Sections 1.3.3 and 1.3.4 do the same for the monetary rule set included in this model. Sections 1.3.5 , 1.3.6 and 1.3.7 complete the model setup by describing the household and firm problem followed by closing conditions. Lastly, section 1.3.8 describes the prior distribution of the model parameters.

1.3.1 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.[3] 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 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

[3] 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.

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

1.3.2 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.

1.3.3 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.

1.3.4 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.

1.3.5 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:

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

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

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

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

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, $\varXi_t(i)$. 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:

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

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

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

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.

1.3.6 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:

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

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.

1.3.7 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:

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

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

1.3.8 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 1.1 and Table 1.2.

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 1.1: 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).[4] 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 already features a large number of parameters and requires a significant amount of computational resources.

[4] 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 1.2: 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.

1.4 Estimation procedure

1.4.1 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. 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 paper, 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}:

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

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:

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

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

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

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.

1.4.2 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 paper.

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.

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. $Initialization$:

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

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

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:

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

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 here.

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*:

$while\ \phi_n < 1$

$set\ n = n + 1\ and\ choose\ \phi_n = g()$

$correction\ step$:

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

$selection\ step$:

$$ESS_n = \frac{N}{\frac{1}{N}\sum_{i=1}^N (\widetilde{W}_n^i)^2}$$

$if\ ESS_n \geq 0.5 \times N$

$W_n^i = W_n^i\ \ and\ \theta_n^i = \theta_{n-1}^i\ \forall\ i \in \{1, \dots, N\}$

$else$

$use\ systematic\ resampling\ with\ \{\theta_{n-1}^i, W_n^i\}_{i=1}^N$

$generate\ sample\ \ \{\theta_n^i\}_{i=1}^N with\ W_n^i = 1$

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

1.4.3 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 paper 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.

1.4.4 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 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 I utilize 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 here 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 $\frac{n_{eval} \times t_{likelihood\ eval}}{n_{cores}}$.[5]

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

[5] 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.

1.4.5 Posterior estimates

The following section presents posterior estimates for the model parameters. The results are summarised in Table 1.3, Table 1.4 and Table 1.5.

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 1.3: 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 1.4: 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 1.5.5 traces out the gradients across time to interpret the parameters in a joint fashion.

Table 1.5: 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.

1.5 Fiscal policy effectiveness

1.5.1 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_{t-1}^L.

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_t^Q, is governed by an auxiliary linear system:

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

$$x_t^Q = 0.5 * h_{\sigma\sigma} + H_x x_{t-1}^Q + 0.5 * H_{xx}(x_{t-1}^L \bigotimes x_{t-1}^L) + \sigma J v_t, \qquad v_t \sim N(0, I).$$

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, $z_t = [x_t^{L}{}', x_t^{Q}{}', (x_t^L \bigotimes x_t^L)']'$ 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:

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

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_{t-1}^L.

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

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

where γ is a $(1 \times (2n_x + n_x^2))$ row vector with all elements equal to zero but entry i. $IRF_{t+h}^i(x_{t-1}^L)$ is affine in x_{t-1}^L 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 $IRF_{t+h}^i(x)$ is affine if and only if the following holds:

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

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:

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

Thus, this leads us to conclude that IRF_{t+h}^i 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_{t+h}^i is of the above design, then mechanically, it is feasible for the IRF_{t+h}^i 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 paper, 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.

1.5.2 Impulse response functions at and around the steady state

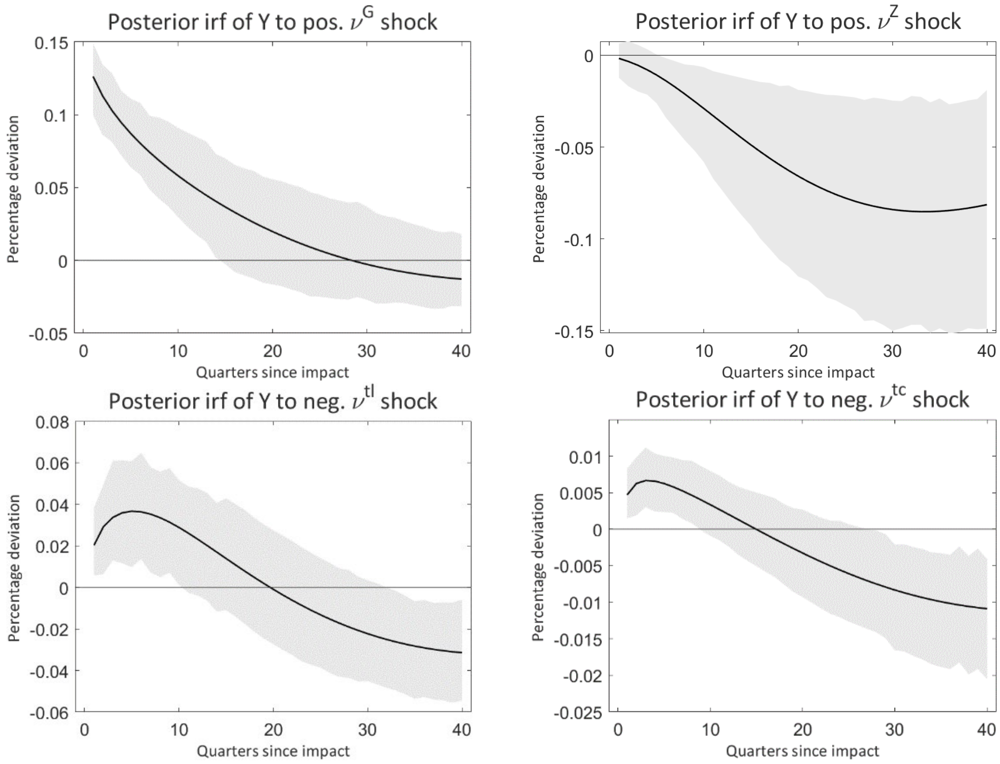
The section analyses the business cycle dependency of impulse responses to fiscal policy shocks. First, I look at the impulse responses of output in the linear states of the DSGE model, and I focus on labour taxation, consumption taxation, government consumption and transfer shocks. As a second step, 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.[6] 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. 1.1 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. 1.1: 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.

[6] 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, 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:

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

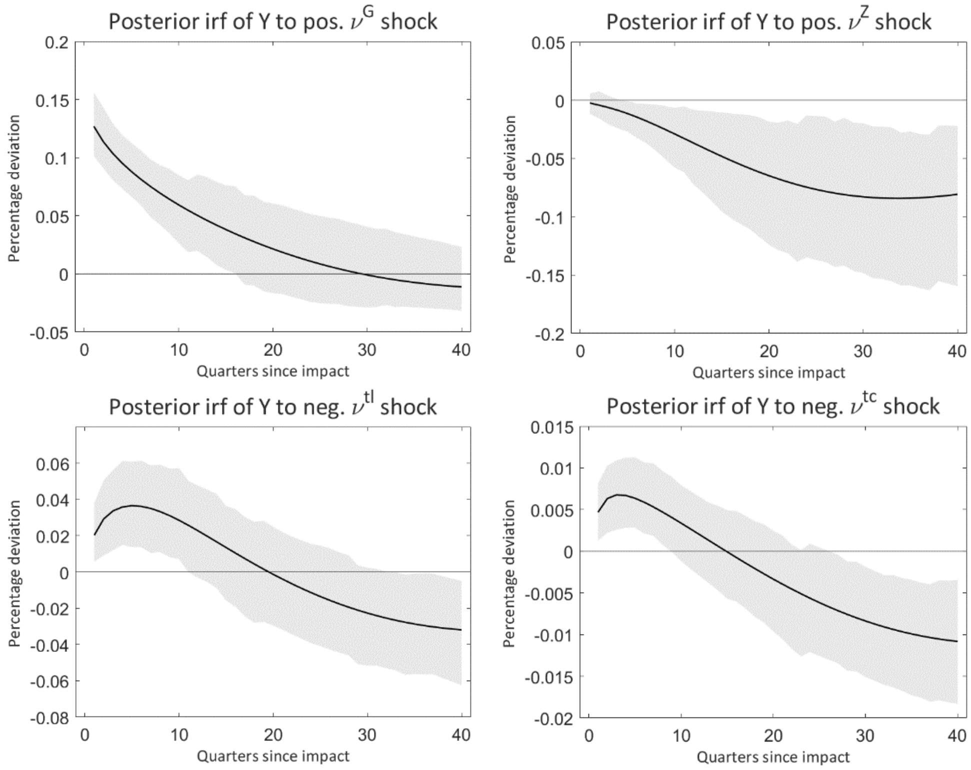
$$\zeta_t = \begin{bmatrix} v_t \\ (v_t \otimes v_t) - vec\big(I_{n_v}\big) \\ (v_t \otimes x^L_{t-1}) \\ (x^L_{t-1} \otimes v_t) \end{bmatrix}.$$

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:

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

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. 1.2, 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. 1.2: non-linear output impulse response to fiscal shocks at steady state

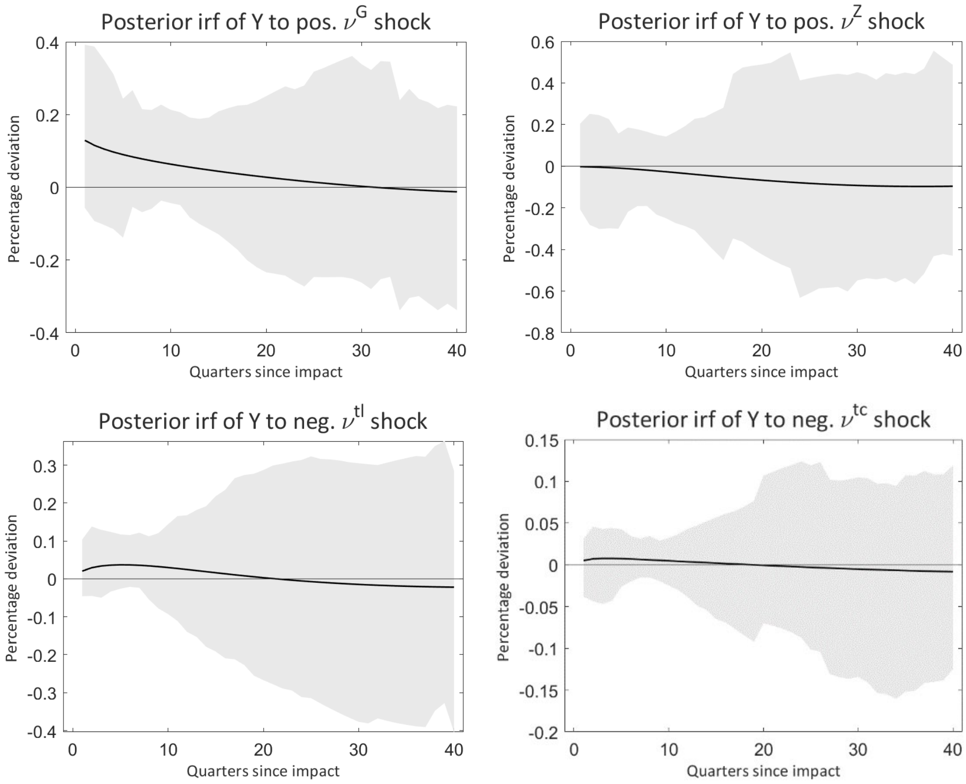


Notes: Impulses responses of non-linear output to one standard deviation fiscal shocks evaluated at the steady state. The solid line is the mean impulse response; the grey shaded is the 95% highest posterior density interval.

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. 1.3. 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. 1.3: 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.

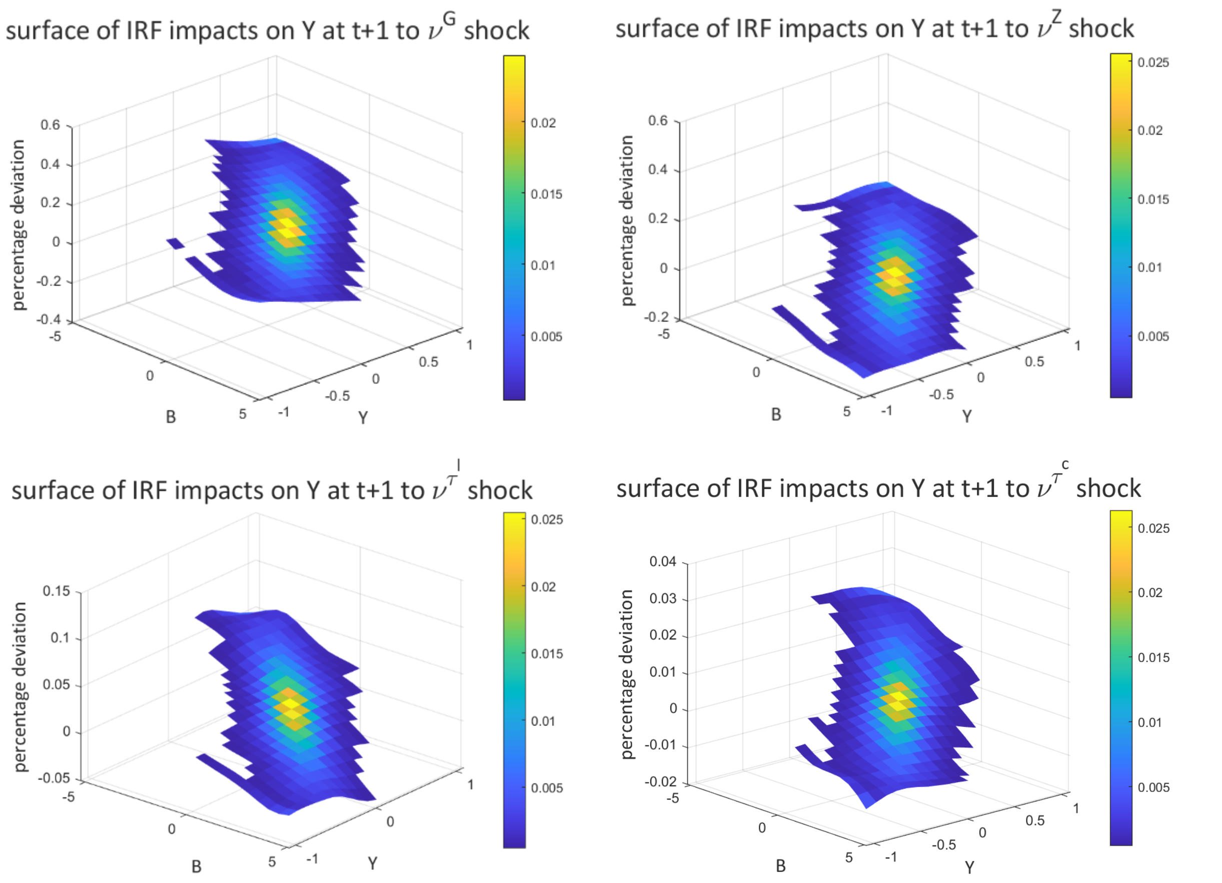
To sum up, the estimated model shows standard and relatively precise impulse responses to fiscal shocks in the linear framework. Changing to the non-linear framework but conducting fiscal policy at the steady state does not offer additional behaviour or conclusions. However, viewing impulse responses 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.

1.5.3 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. 1.4: 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. 1.4 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 1.5.1 , 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.[7] The results are presented in Table 1.6 and Table 1.7. Further tables for IRF_{t+4}^{Y} on initial conditions are provided in the appendix.

Table 1.6: Regression of IRFs on impact of output to gov. consumption and transfer shocks on initial conditions

variable	$IRF_{t+1}^{Y} \mid v^{G}$			$IRF_{t+1}^{Y} \mid v^{Z}$		
	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		

[7] 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 1.7: 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.[8]

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.

1.5.4 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.

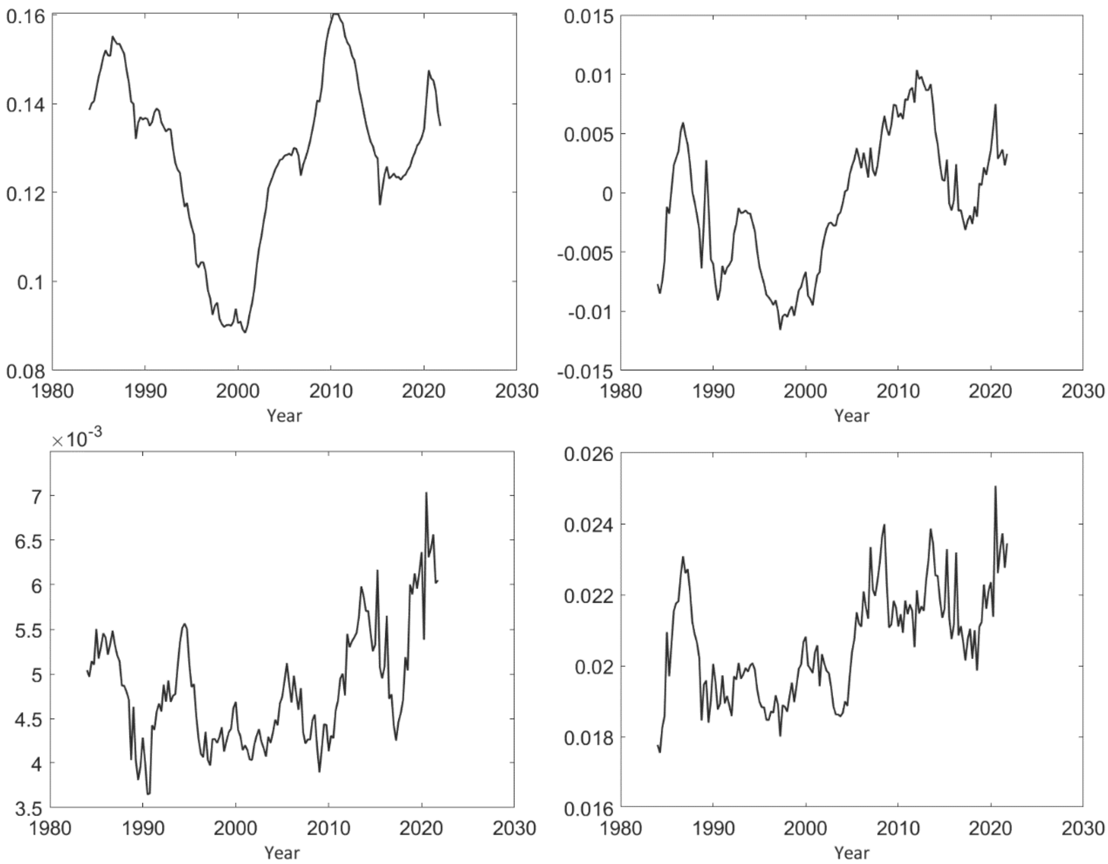
To do so, I construct the posterior mean estimates of the state vectors across the sample based on the conditional particle filter using parameter draws from the posterior. That includes estimates for all state variables of the DSGE, including inflation, output, interest rate and more. These state vectors 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. 1.5 and Fig. 1.6 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,

[8] 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. 1.5: Paths of impact effect of fiscal policy around the cycle

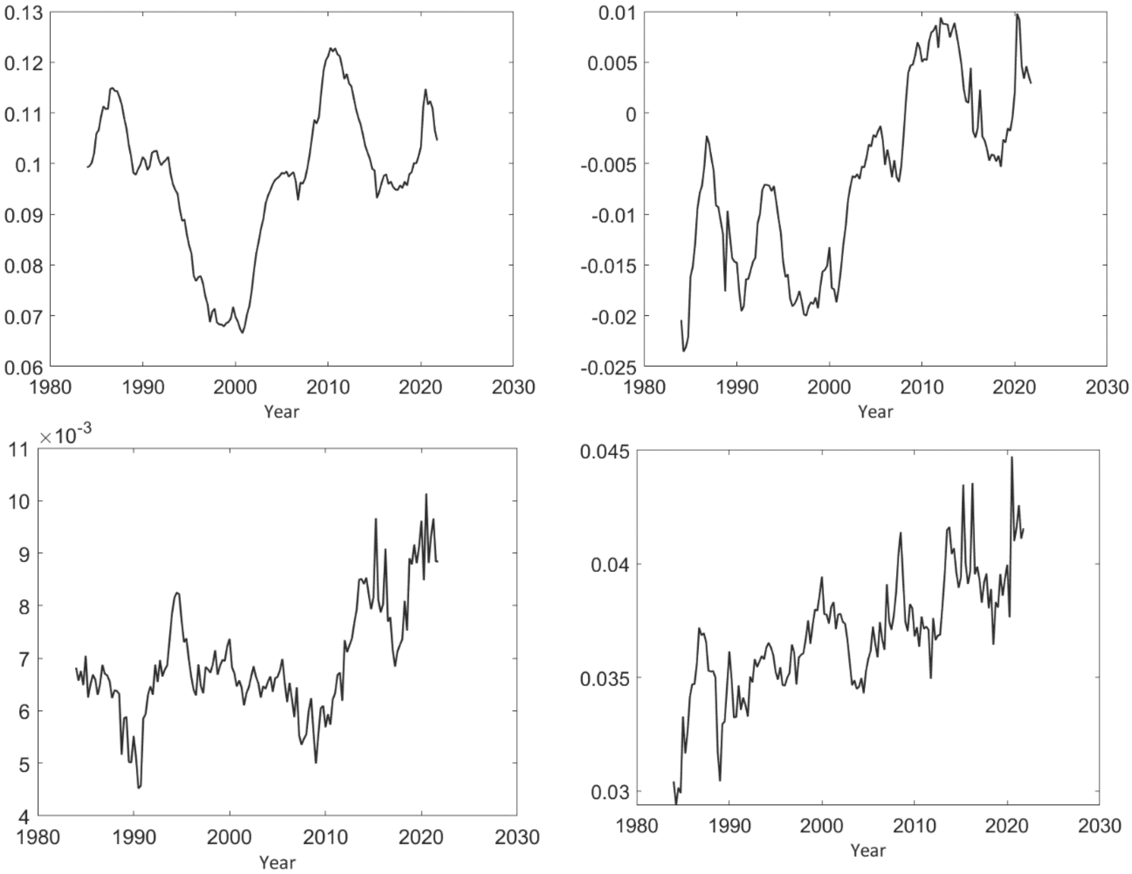


Notes: Constructed paths for IRF^{Y}_{t+1} in response to fiscal shocks for government consumption (upper left), transfers (upper right), consumption taxation (lower left) and labour taxation (lower right)

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. 1.5 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 1.5.5 , 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. 1.6: Paths of effect of fiscal policy at four quarters around the cycle



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

1.5.5 Policy gradients

The policy rules for the federal government and the central bank 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:

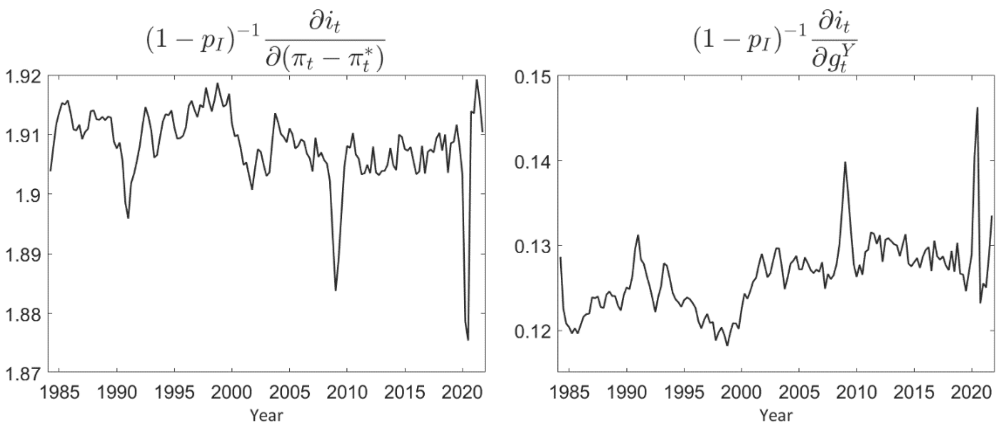
$$\psi_{y,t} = (1-\rho_I)^{-1}\frac{\partial i_t}{\partial(y_t - y_{t-1})} = \big(\psi_y + \psi_{y,y}(y_t - y_{t-1}) + \psi_{y,\pi}(\pi_t - \pi_t^*)\big),$$

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

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. 1.7.

Fig. 1.7: 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.

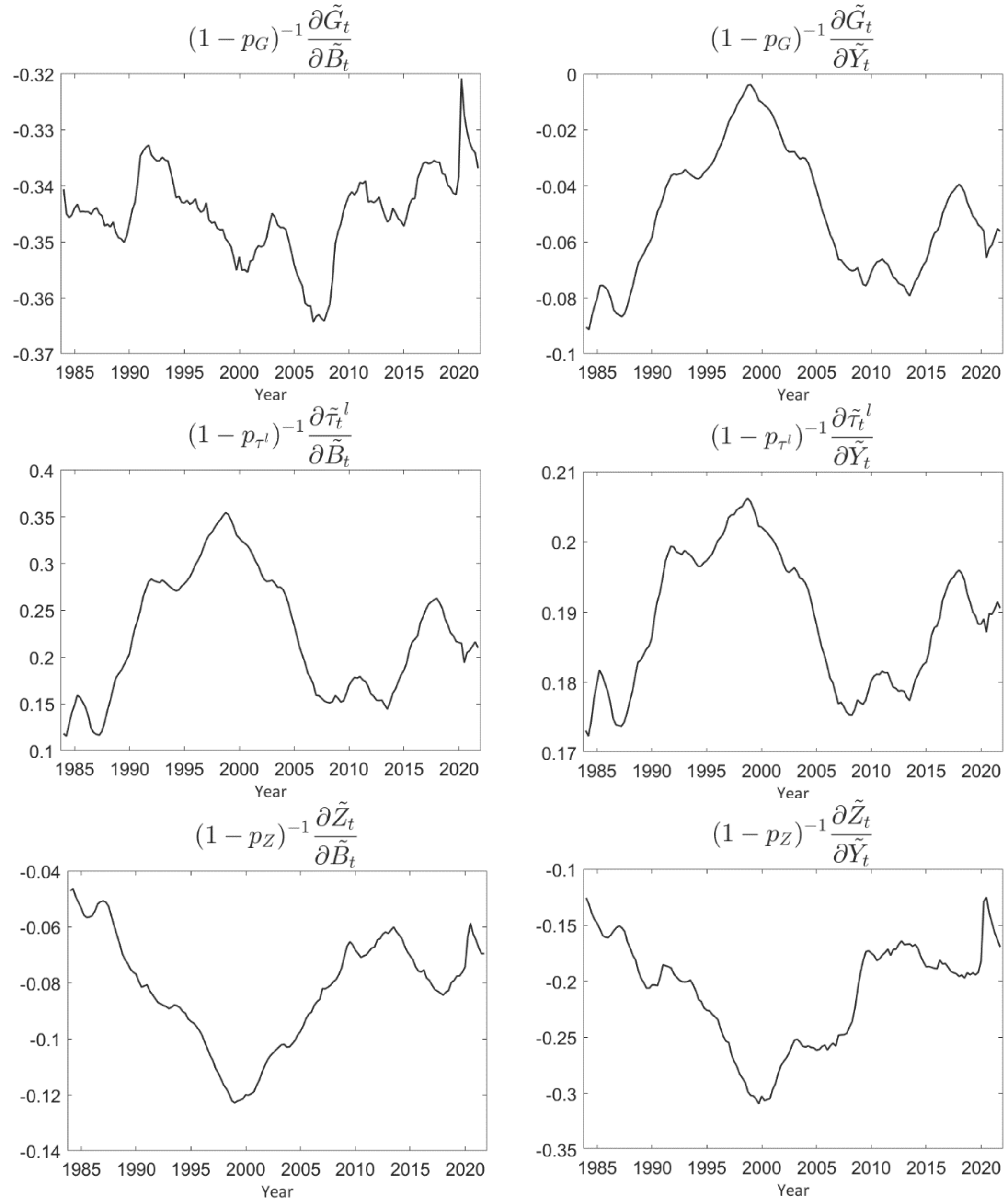
However, the persistence, scale and recovery of the gradient changes seem to differ from crisis to crisis. Mechanically, the reason for this is that the gradients are highly correlated with output growth. $\psi_{\pi,t}$ is positively correlated to output growth and $\psi_{y,t}$ negatively. Consequently, 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. 1.7 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. 1.7 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. 1.8. 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. 1.8: 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 $\frac{\partial \tilde{Z}_t}{\partial \tilde{Y}_t}$.

1.6 Conclusion

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 paper 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 paperr 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 paper 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.

1.7 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).

Appendix A

A.1 Second-order pruned system

For the estimations in this paper, 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 $z_t = [x_t^{L\prime}, x_t^{Q\prime},\ (x_t^{L} \bigotimes x_t^{L})']$ 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:

$$
\begin{bmatrix} x^L_{t+1} \\ x^Q_{t+1} \\ (x^L_{t+1} \bigotimes x^L_{t+1}) \end{bmatrix}
$$

$$
= \begin{bmatrix} 0 \\ 0.5 * h_{\sigma\sigma} \\ \sigma^2 (J \bigotimes J) vec\big(I_{n_v}\big) \end{bmatrix} + \begin{bmatrix} H_x & 0 & 0 \\ 0 & H_x & H_{xx} \\ 0 & 0 & (H_x \bigotimes H_x) \end{bmatrix} \begin{bmatrix} x^L_t \\ x^Q_t \\ (x^L_t \bigotimes x^L_t) \end{bmatrix}
$$

$$
+ \begin{bmatrix} \sigma J & 0 & 0 & 0 \\ \sigma J & 0 & 0 & 0 \\ 0 & \sigma^2 (J \bigotimes J) & \sigma^2 (J \bigotimes H_x) & \sigma^2 (H_x \bigotimes J) \end{bmatrix} \begin{bmatrix} v_{t+1} \\ (v_{t+1} \bigotimes v_{t+1}) - vec\big(I_{n_v}\big) \\ (v_{t+1} \bigotimes x^L_t) \\ (x^L_t \bigotimes v_{t+1}) \end{bmatrix},
$$

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:

$$
y_{t+1} = d_2 + C_2 z_{t+1},
$$

$$
y_{t+1} = [0.5 * g_{\sigma\sigma}] + [0 \quad G_x \quad 0.5 * G_{xx}] \begin{bmatrix} x^L_{t+1} \\ x^Q_{t+1} \\ (x^L_{t+1} \bigotimes x^L_{t+1}) \end{bmatrix}.
$$

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}.

A.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 paper, 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.[9] 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

[9] 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 $\mathit{repelem}(X_t^{system}, n_x, 1) .* \mathit{repmat}(X_t^{system} n_x, 1)$, where ".*" indicates element by element multiplication. The last step is to pre-multiply by H_{xx}.

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.

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 A.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 A.1: Overview of performance gains across different specifications

Notes: Time is measured in seconds

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

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.[10] 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.

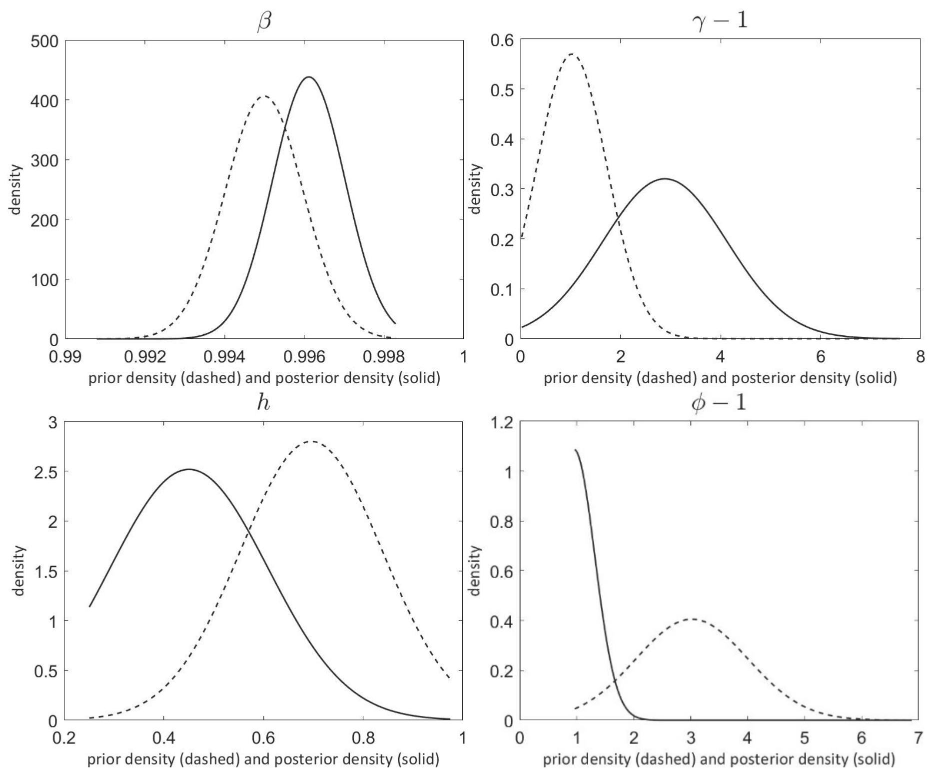
[10] The actual speed up depends on the exact algorithm settings (comparing reasonable algorithm settings vs forcing equal numbers of likelihood evaluations) but is substantial in any case. For an estimation using the Metropolis Hastings algorithm using a quite standard 5,000,000 likelihood evaluations 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%).

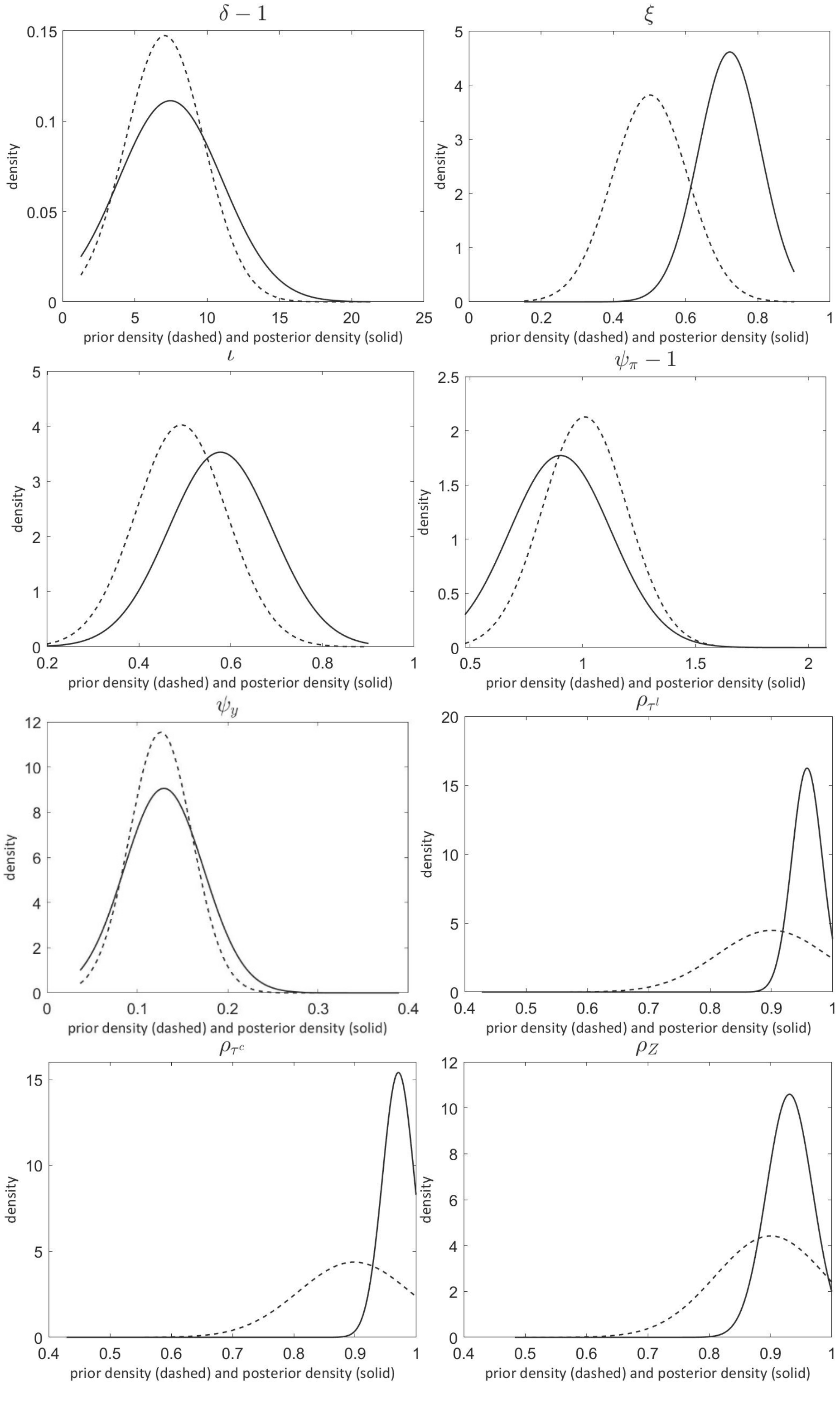
For this paper, I am really grateful to Amisano and Tristani (2010) for providing me with their code base, which allowed me to double-check my work and improve upon it.

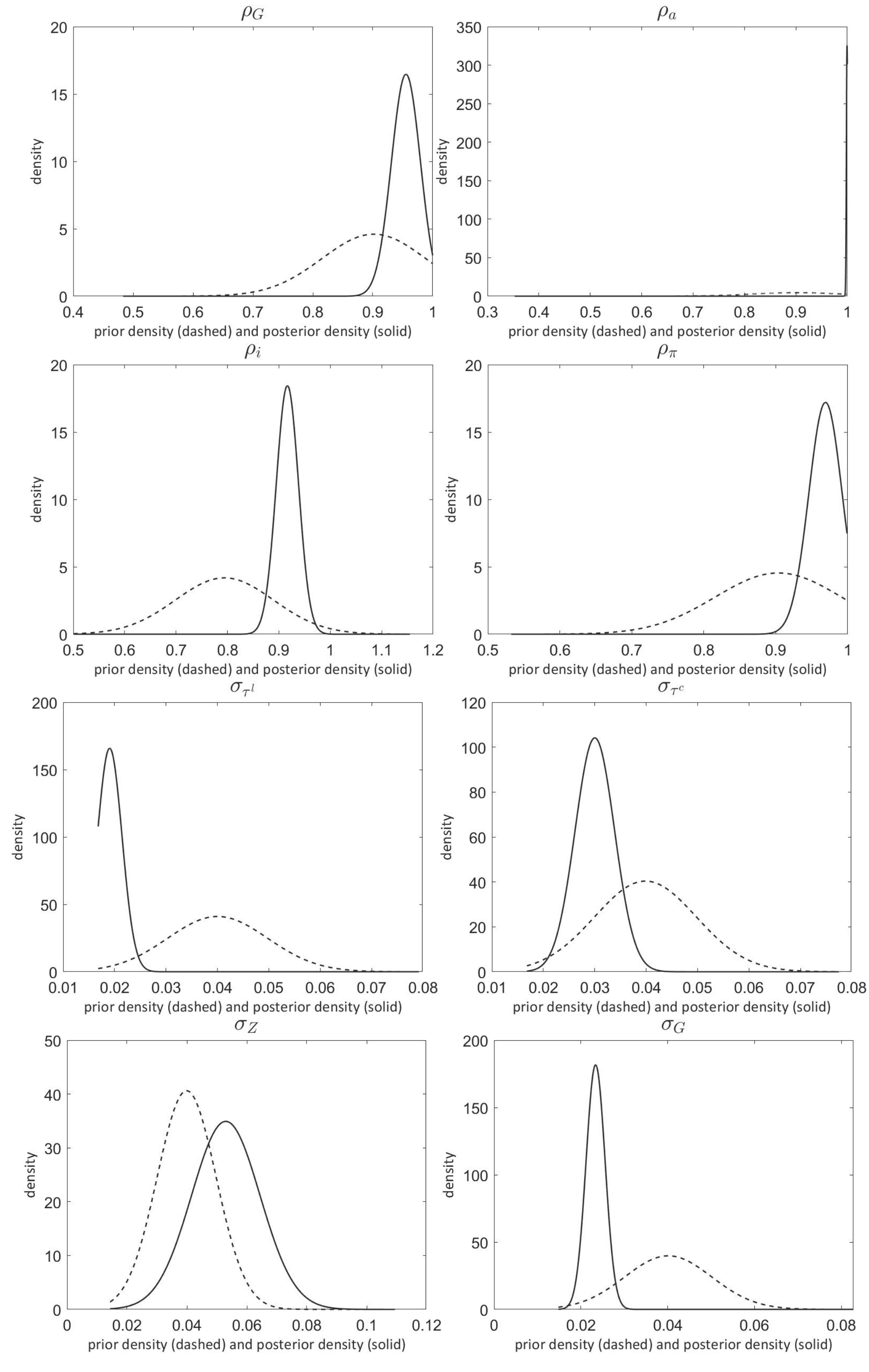
A.3 Posterior density plots

A.3.1 Core model parameters

Fig. A.1: Posterior density graphs for core model parameters



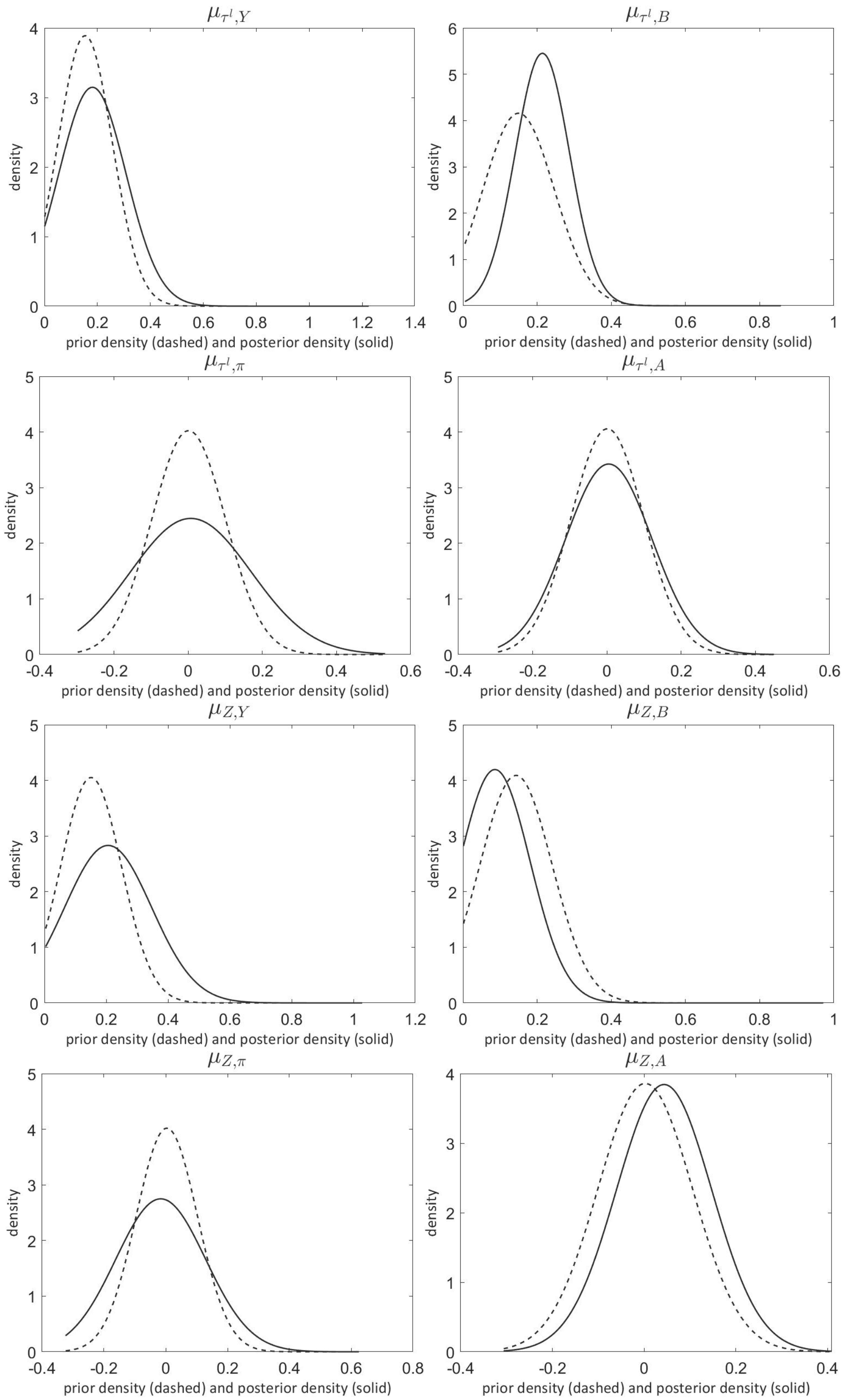

$\delta - 1$
ξ
ι
$\psi_\pi - 1$
ψ_y
ρ_{τ^l}
ρ_{τ^c}
ρ_Z
density
prior density (dashed) and posterior density (solid)

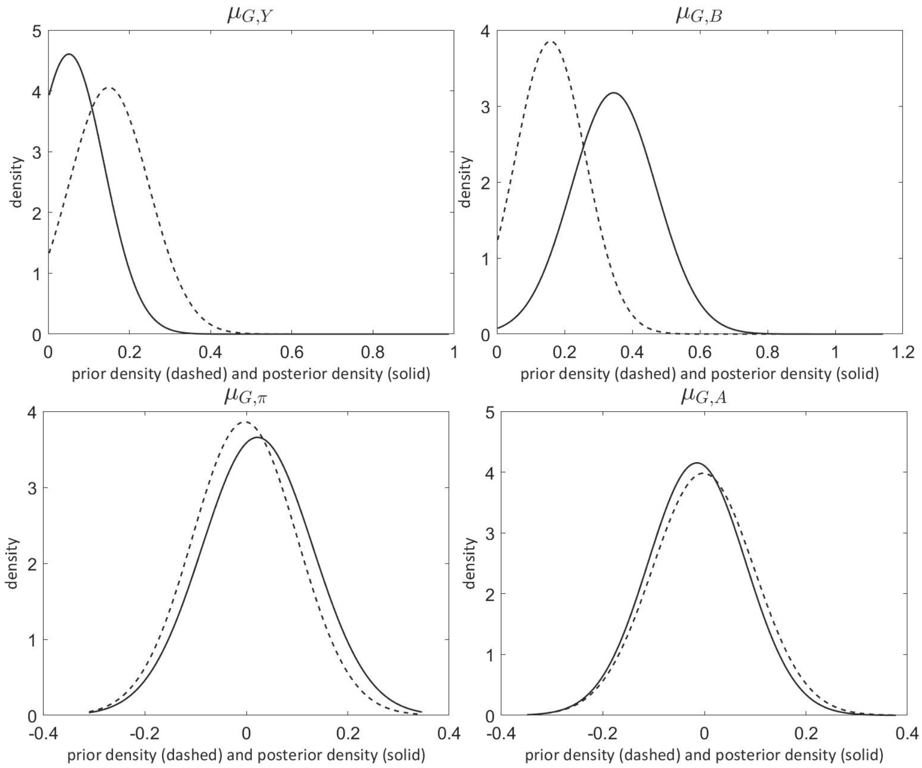


Notes: Posterior and prior density plots for core model parameters. Dashed line corresponds to prior density and solid line to posterior density.

A.3.2 Linear rule parameters

Fig. A.2: Posterior density graphs for linear rule parameters

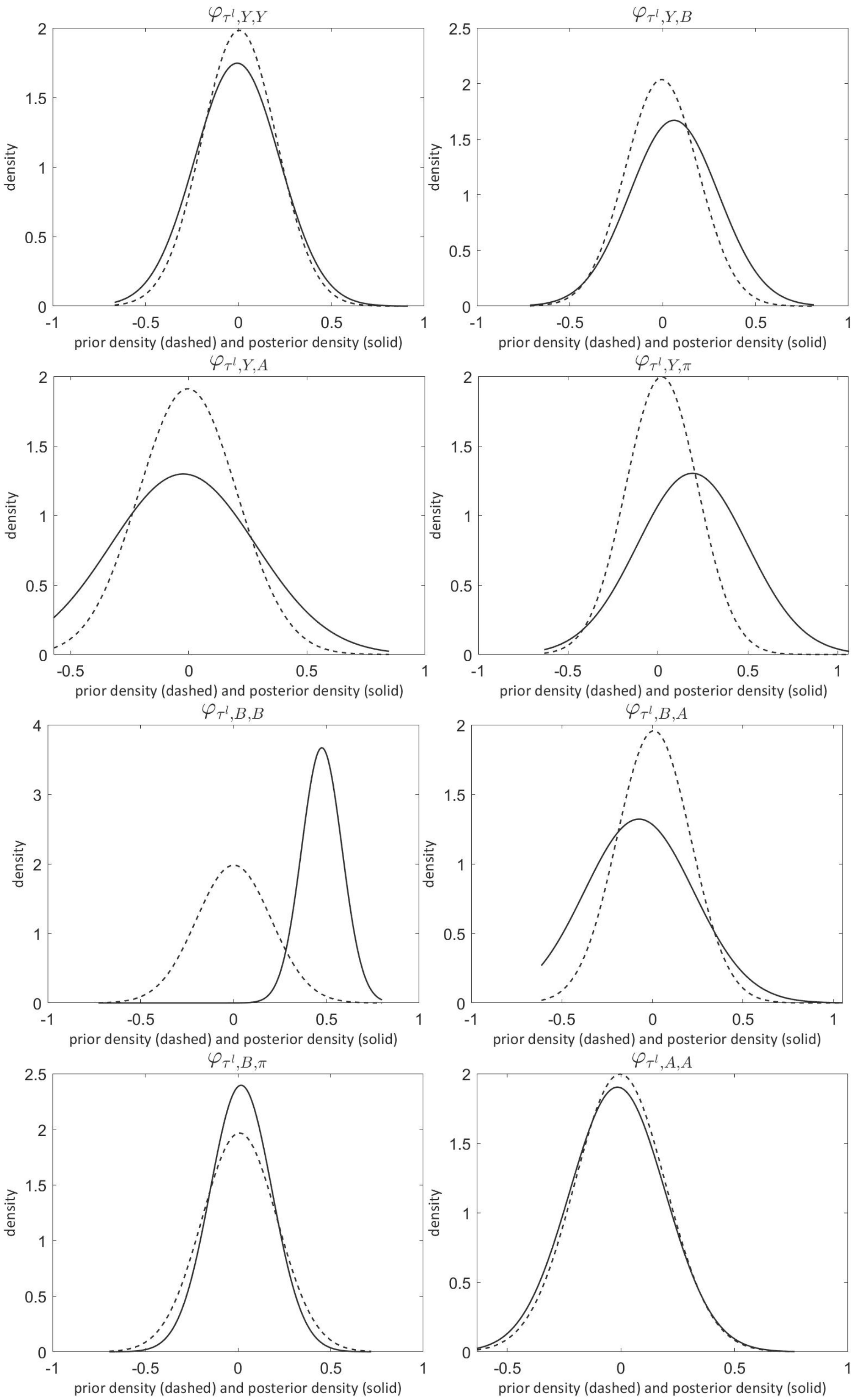


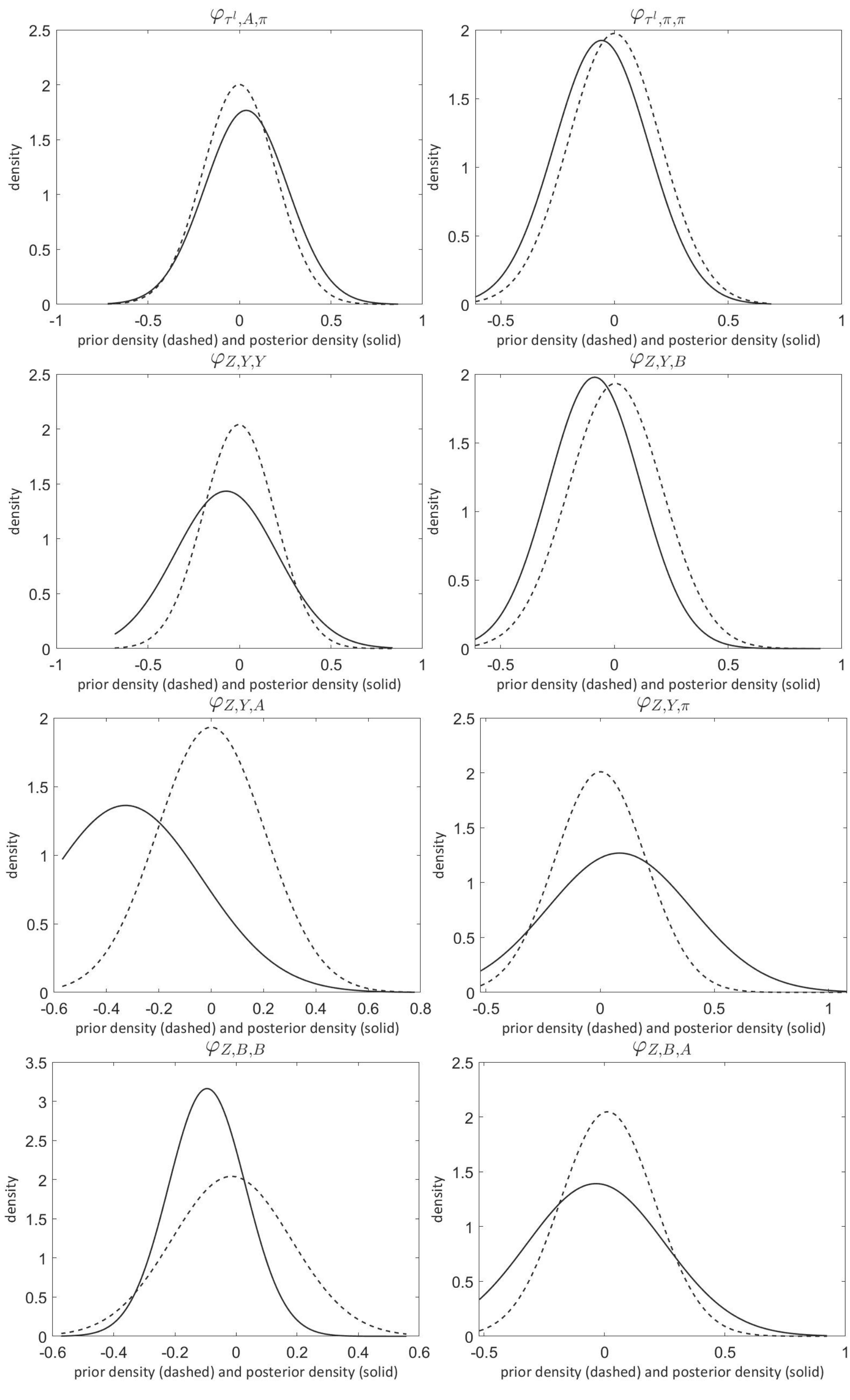


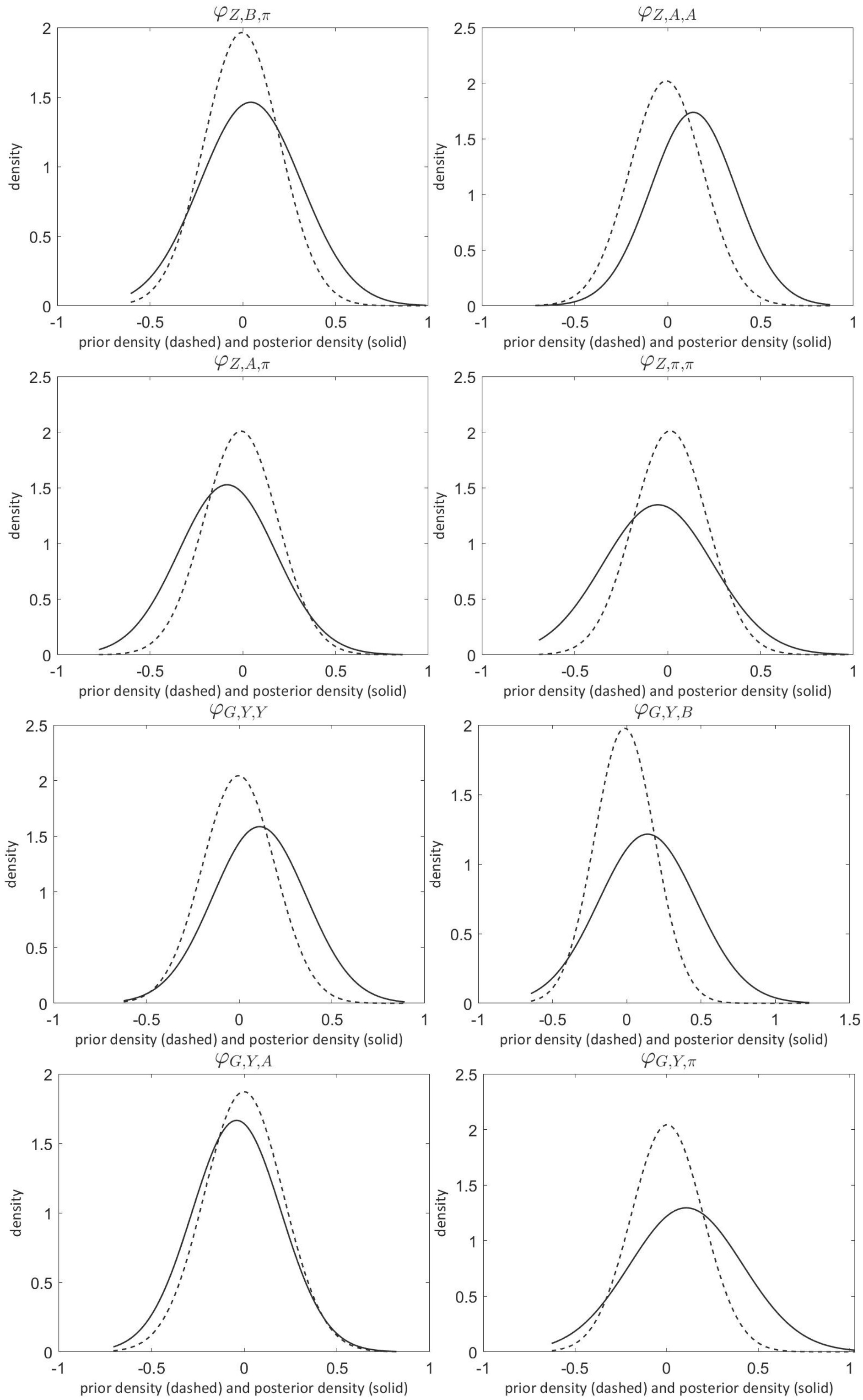
Notes: Posterior and prior density plots for linear fiscal rule parameters. Dashed line corresponds to prior density and solid line to posterior density.

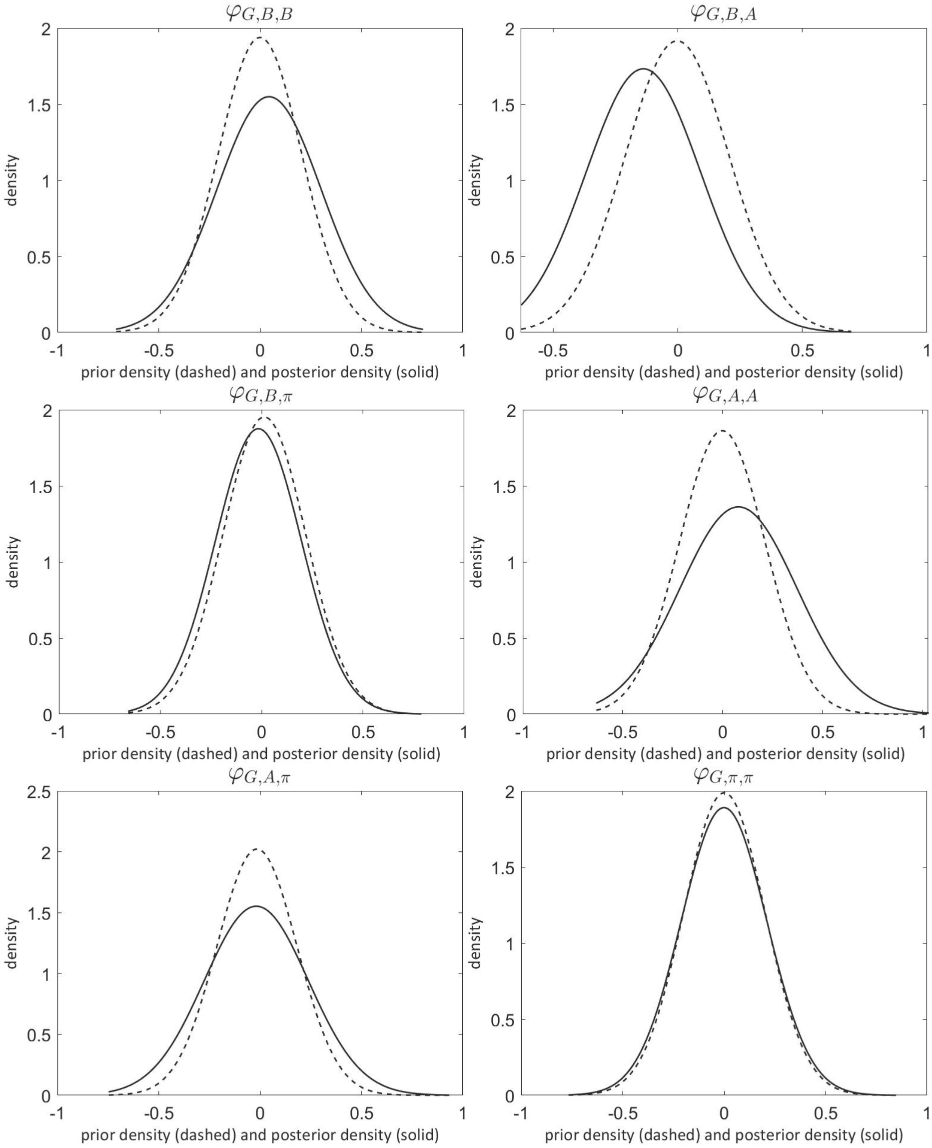
A.3.3 Non-linear rule parameters

Fig. A.3: Posterior density graphs for non-linear rule parameters

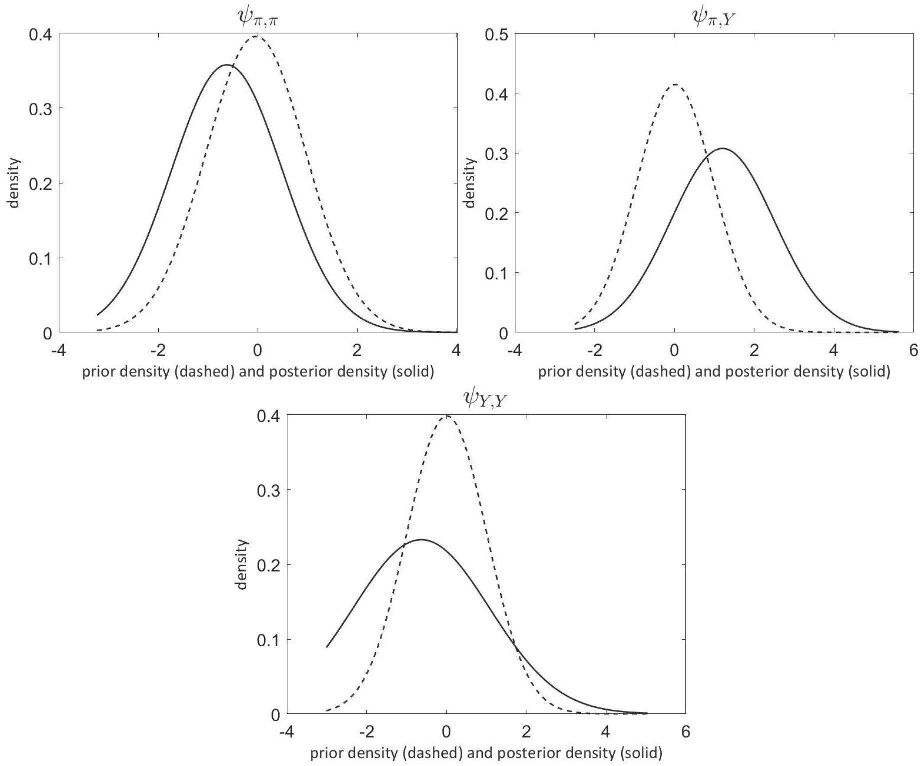



$\varphi_{\tau^l,A,\pi}$
$\varphi_{\tau^l,\pi,\pi}$
$\varphi_{Z,Y,Y}$
$\varphi_{Z,Y,B}$
$\varphi_{Z,Y,A}$
$\varphi_{Z,Y,\pi}$
$\varphi_{Z,B,B}$
$\varphi_{Z,B,A}$
density
prior density (dashed) and posterior density (solid)


$\varphi_{Z,B,\pi}$
$\varphi_{Z,A,A}$
$\varphi_{Z,A,\pi}$
$\varphi_{Z,\pi,\pi}$
$\varphi_{G,Y,Y}$
$\varphi_{G,Y,B}$
$\varphi_{G,Y,A}$
$\varphi_{G,Y,\pi}$
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.

A.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 here 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 paper. 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 A.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 A.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 A.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 A.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

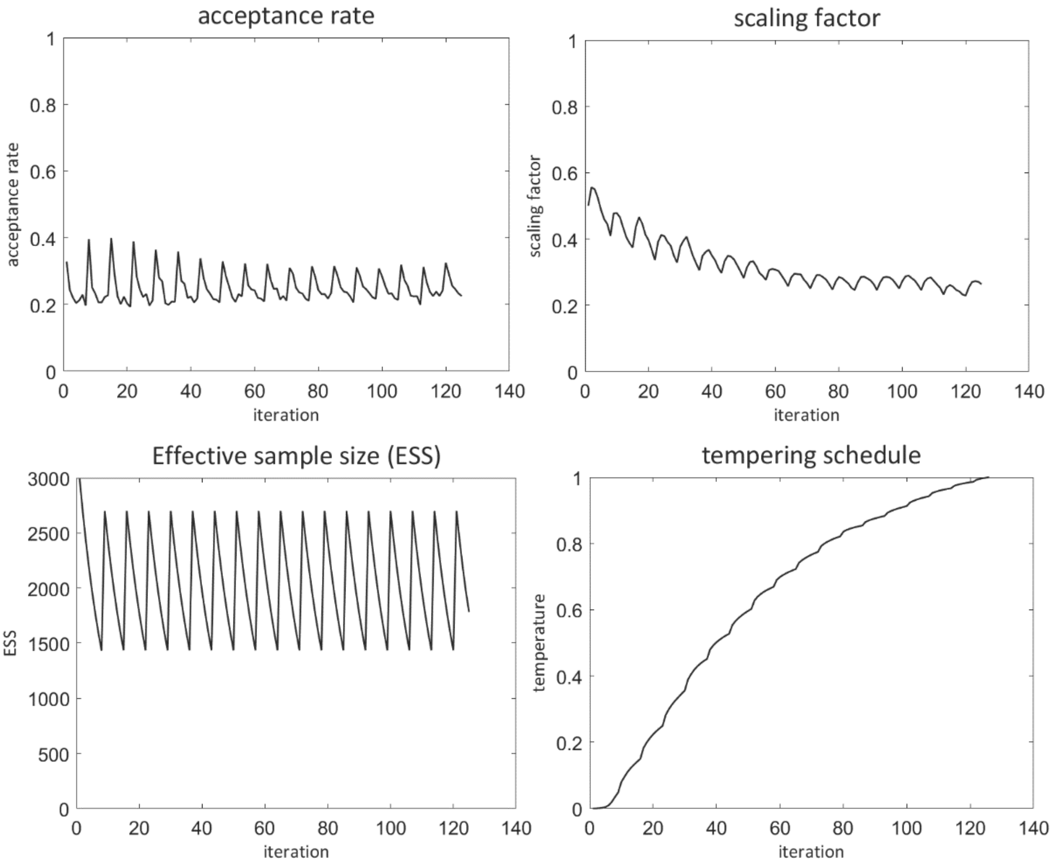
A.5 Estimation diagnostics

Fig. A.4 below displays estimation diagnostics for the main estimation in this paper. 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.[11] 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.

[11] 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. A.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.

A.6 Regression Tables for IRFs

Table A.4: Regression of IRFs on impact of output to government consumption and transfer shocks on initial conditions

variable		$IRF^Y_{t+4}\|\upsilon^G$			$IRF^Y_{t+4}\|\upsilon^Z$	
	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
$\tilde{\iota}_{t-1}$	-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
$\tilde{\tau}^l_{t-1}$	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
$\tilde{\tau}^c_{t-1}$	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^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 A.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.