European University Institute

DEPARTMENT OF ECONOMICS

EUI Working Papers

ECO 2009/29
DEPARTMENT OF ECONOMICS

MATCHING THEORY AND DATA: BAYESIAN VECTOR AUTOREGRESSION AND DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODELS

Alexander Kriwoluzky

EUROPEAN UNIVERSITY INSTITUTE, FLORENCE
DEPARTMENT OF ECONOMICS

Matching Theory and Data: Bayesian Vector Autoregression and Dynamic Stochastic General Equilibrium Models

ALEXANDER KRIWOLUZKY

EUI Working Paper **ECO** 2009/29

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Matching Theory and Data: Bayesian Vector Autoregression and Dynamic Stochastic General Equilibrium Models*

Alexander Kriwoluzky[†]
European University Institute Florence

June 29, 2009

Abstract

This paper shows how to identify the structural shocks of a Vector Autoregression (VAR) while simultaneously estimating a dynamic stochastic general equilibrium (DSGE) model that is not assumed to replicate the data-generating process. It proposes a framework for estimating the parameters of the VAR model and the DSGE model jointly: the VAR model is identified by sign restrictions derived from the DSGE model; the DSGE model is estimated by matching the corresponding impulse response functions.

JEL classification: C51.

Keywords: Bayesian Model Estimation, Vector Autoregression, Identification.

*I am especially thankful to Harald Uhlig, Chris Sims, Bartosz Mackowiak, Helmut Lütkepohl, Wouter den Haan, Francesco Ravazzolo, Lenno Uusküla, Morten Ravn and Martin Kliem for their comments and suggestions. The paper further benefitted from seminar participants at Norges Bank. Part of the research was conducted while I was visiting Princeton University funded by the German Academic Exchange Service. Further grants from the DEKA Bank and the SFB 649 are gratefully acknowledged.

[†]European University Institute, Max Weber Programme, Via delle Fontanelle, 10, 50014 San Domenico, Italy, email: Alexander.Kriwoluzky@eui.eu, tel: +39 055 4685 851.

1 Introduction

How can we estimate the effects of an exogenous disturbance on the economy? In recent years, two methodologies have become popular to answer this question: the Vector Autoregression (VAR) model and the Dynamic Stochastic General Equilibrium (DSGE) model approach. Both have considerable advantages but also substantial shortcomings. While on the one hand the VAR model is straightforward to estimate, structural shocks cannot be recovered without additional assumptions. The DSGE model, on the other hand, is of a structural form, i.e. it exhibits structural shocks, but it is difficult to determine its parameter values. In this paper I present a methodology for estimating the effects of exogenous disturbances that combines the advantages of both approaches while overcoming their respective limitations.

I suggest identifying the VAR model with the help of the structural impulse response functions of the DSGE model. Deriving the identifying restrictions from the DSGE model ensures consistency of the identification of the VAR model with the implied structural form of the DSGE model. Moreover, this approach allows the researcher to lay out the assumptions underlying the identification of the VAR model explicitly in the DSGE model and it enables her to include the different assumptions she wants to discriminate between in the DSGE model. In this case a larger class of identifying assumptions can be considered a priori and evaluated afterwards. At the same time, the parameters of the DSGE model are estimated using information from the VAR model. This has the advantage that the DSGE model does not have to be assumed to represent the data-generating process nor to be fully stochastically specified. Therefore, it need not exhibit as many structural shocks as there are observable variables to be explained. Moreover, features and frictions which are not pertinent to the question being examined can be ignored.

More precisely, the VAR model is identified using sign restrictions derived from the structural impulse response functions of the DSGE model, while the DSGE model is estimated by matching the corresponding impulse response functions. Transferring the restrictions via sign restrictions is straightforward and easy to handle: for a given parametrization of the DSGE model the signs of the impulse response functions of the DSGE model define the restrictions for identifying the VAR model. Furthermore, when using sign restriction it is not necessary for the complete number of structural shocks of the VAR model to be identified, nor need the number of structural shocks of the DSGE model correspond to the number of observable variables (variables in the VAR model). The parameter vector of the DSGE model is in turn estimated by matching the corresponding impulse response functions of the VAR model. Thus, it only needs to represent the dynamics of the economy, not the complete data-generating process. Consequently, features and lags which would otherwise have been included to match outliers in the data, but which are not essential to the study, can be dropped.

In order to carry out this estimation procedure, it is necessary to describe the joint distribution of the VAR model and the DSGE model. This paper presents a methodology for doing so. The methodology is first illustrated by means of a Monte Carlo experiment and then applied to the data. I employ two different DSGE models in each exercise. This is motivated by the fact that the simple DSGE model used in the Monte Carlo Experiment exhibits different signs in the response of each variable depending on the parametrization, i.e. it is a perfect example, but is too stylized to be estimated. The DSGE model used in the estimation exercise does not exhibit this characteristic. Only the response of one variable, the one under investigation, switches signs across the parameter space. However, it is straightforward to be taken to the data.

More precisely, I simulate data from a fiscal theory of the price level (FTPL) model and re-estimate the parameters of the FTPL model and the impulse response functions of the VAR model. The experiment shows that the true impulse response function is indeed found. The FTPL model serves well for illustrating purposes since it can be reduced to two equations in two variables and two shocks. The signs of both variables vary depending on two parameters only. It is less well suited to bringing it to the data. I therefore estimate a DSGE model recently laid out by Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008) to investigate the response of inflation to a monetary policy shock. This DSGE model suits well, since the response of inflation is either positive or negative depending on its parametrization.

The paper is organized as follows: The next section briefly reviews the relevant literature. The third section outlines the general framework. The fourth section describes the probability distributions and the algorithm suggested to approximate them in detail. The Monte Carlo Experiment is conducted in section 5. Section 6 applies the methodology to the data and estimates the deep habit model. The last section concludes.

2 Related Literature

After Sims's seminal article (Sims, 1980) VAR models became one of the workhorses in macroeconomics despite the problem of identifying structural shocks. Suggestions for resolving the identification problem in a VAR model are manifold. Excellent surveys have been written by Christiano, Eichenbaum, and Evans (1999) and Rubio-Ramirez, Waggoner, and Zha (2005). The approaches most closely related to the methodology presented here are to identify the VAR model by sign restrictions (Uhlig, 2005a; Faust, 1998) or by probabilistic restrictions (Kociecki, 2005). Identification employing sign restrictions attempts to restrict the signs of the impulse response functions of some variables, while the variable of interest is unrestricted. In Kociecki (2005), a prior distribution for the impulse response functions is formulated and transformed into a prior distribution for the coefficients of the structural VAR model. Both approaches depend on the availability of a priori knowledge on the behavior of some impulse response functions.

With regard to explicitly basing the identifying assumptions on DSGE models, two strands of literature have emerged recently. One derives the identifying assumptions from a DSGE model (Altig, Christiano, Eichenbaum, and Linde (2002), DelNegro and Schorfheide (2004) and Sims (2006b)); the other suggests, once the DSGE model is large enough, estimating the DSGE model and thereby directly infering on the impulse responses (as in Smets and Wouters (2003) and Smets and Wouters (2007)).

Due to advances in computational power, the estimation of DSGE models has lately become very popular. The procedures differ depending on the econometric interpretation of the DSGE model. Geweke (1999) distinguishes between a strong and weak interpretation. The former requires the DSGE model to provide a full description of the data-generating process. It is the more common one nowadays despite its shortcomings: first, the DSGE model already puts a lot of structure on the impulse responses a priori, i.e. it often does not allow an investigation of the sign of a response and might therefore not be appropriate as a research tool. Second, not all parameters of the DSGE model can be identified (see Canova and Sala (2006) and Beyer and Farmer (2006)). Finally, not all economists might feel comfortable with the assumption that the DSGE model is a proper representation of the data-generating process. Instead, as mentioned in Christiano, Eichenbaum, and Evans (2005), the DSGE model is best suited to replicate the implied dynamics in the data, i.e. the impulse response functions. This is the weak econometric interpretation. Following this road Ravn, Schmitt-Grohé, and Uribe (2007), Mertens and Ravn (2008) as well as Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008) estimate a DSGE model given the impulse response function of the VAR model by minimizing the distance between the corresponding impulse response functions. In contrast to them I do not consider the impulse response functions of the VAR model as given, i.e. as identified a priori. In the case of timing or long-run restriction the VAR model is identified and considering the impulse response functions as given is justified. This paper addresses the cases when the identifying restrictions are not a priori clear or when the researcher chooses to use sign restrictions. Sign restrictions derived from a DSGE model will only in very rare cases be unique across the parameter space of the DSGE model. In those cases the impulse response functions are not identified and one cannot proceed as for instance in Ravn, Schmitt-Grohé, and Uribe (2007), Mertens and Ravn (2008) or Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008).

The methodology presented in this paper is in the spirit of the former strand of the literature, i.e. it bases the identification of the VAR model on restrictions derived from the DSGE model. It differs from the existing literature in the following aspects. Altig, Christiano, Eichenbaum, and Linde (2002) and DelNegro and Schorfheide (2004) employ the rotation matrix of the DSGE model to identify the VAR model. To do this, the DSGE model has to be fully stochastically specified. In the case of DelNegro and Schorfheide (2004), additional dummy observations derived from the model are used to augment the VAR model as suggested originally by Ingram and Whiteman (1994). While one can control for the prior weight of the dummy observations, one cannot control for the prior weight of the implied dynamics of the DSGE. The methodology proposed here differs from this by not employing the implied rotation matrix of the DSGE model to identify the VAR model, and therefore not requiring the DSGE model to be fully stochastically specified.

Sims (2006b) extends the idea to augment the VAR model with dummy observations in a more general framework. In his framework, the tightness of the prior can be varied across frequencies and the number of structural shocks does not need to equal the number of observations. The main difference to Sims (2006b) is that I suggest employing the implied sign and shape restrictions (as described in Uhlig (2005a)) to identify the VAR model as it is more simple and straightforward to use.

In recent studies, Lanne and Lütkepohl (2005), Lanne and Lütkepohl (2008), and Lanne, Lütkepohl, and Maciejowska (2009) employ additional statistical properties of the error terms to identify the VAR model. Lanne and Lütkepohl (2005) make use of possible non-normal distributions of the error terms and extract additional identifying information from this. Lanne and Lütkepohl (2008) use the insight of Rigobon (2003) that a VAR model can be identified exploiting changes in volatility. Given any exact identifying scheme this characteristics delivers over-identifying restrictions which can be used to test different identification schemes. In Lanne, Lütkepohl, and Maciejowska (2009) the authors combine the properties of mixed normal distributions and regime changes in the volatility of the error term and show that the VAR model is just identified, given that the shocks are orthogonal across regimes and only the variances of the shocks change across regimes. The methodology presented in this paper does not hinge on special properties of the error terms. It applies also in cases where the residuals are normally distributed.

3 Framework

In this section I set up the VAR model and its corresponding Vector Moving Average (VMA) representation. The issue of how the structural impulse response can be identified is equivalent for both notations. For every period, the impulse response functions of a VAR model can be expressed solely in terms of the coefficients of the VMA model of that period. Setting up the framework in terms of the VMA representation makes the subsequent analytical calculations less demanding. Since the VAR model is connected with the DSGE model via their implied dynamics, the notation necessary for the DSGE model is introduced before the central idea of how to derive the joint posterior distribution for the VAR model and the DSGE model is presented. Afterwards, the framework is related to existing and nested approaches.

3.1 The VAR model and its corresponding VMA model

The structural VAR model containing m variables is defined as:

$$A^{-1}Y_t = A_1Y_{t-1} + A_2Y_{t-2} + \dots A_lY_{t-l} + \epsilon_t, t = 1, \dots, T \tag{1}$$

Y_t is a $m \times 1$ vector at date $t = 1 + l, \dots, T$, A and A_i are coefficient matrices of size $m \times m$ and ϵ an $i.i.d.$ one-step-ahead forecast error, distributed: $\epsilon \sim \mathcal{N}(0, I_{m\times m})$.

The impulse response function φ^V of the VAR model is defined as the response of Y to an innovation in ϵ. Denote the VMA representation as:

$$Y_t = \sum_{i=0}^{\infty} \Theta_i \epsilon_{t-i}, \tag{2}$$

where Θ_i denotes a moving average coefficient matrix. The impulse response function of a VAR model to an innovation in variable i at horizon k φ^V_{jk} can be computed directly as:

$$\varphi^V_{jk} = \Theta_{jk}, \tag{3}$$

where i depicts the i-th column. Due to the assumption that $\Sigma_\epsilon = I_{m\times m}$, this structural moving average representation cannot be estimated directly. Instead the reduced form moving average representation with error term $u_t = A\epsilon_t$, where $u \sim \mathcal{N}(0, \Sigma)$, is estimated. The reduced form moving average coefficients are defined as $\Phi_i = \Theta_i A^{-1}$:

$$Y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i} \tag{4}$$

The factorization $\Sigma = A'A$ does not have a unique solution, which leads to an identification problem of A.

It is important to note that any stationary moving average representation can be approximated by a reduced form VAR model, which takes the form:

$$Y_t = B_1Y_{t-1} + B_2Y_{t-2} + \dots B_lY_{t-l} + u_t, t + 1, \dots, T \tag{5}$$

with $B_i = AA_i$, $u_t = A\epsilon_t$ and $u \sim \mathcal{N}(0, \Sigma)$. While the framework is set up in terms of VMA representation, it can be easily estimated as a VAR model.

3.2 The DSGE model

The fundamental solution of the DSGE model is given by[1]:

$$\hat{x}_t = T(\tilde{\theta})\hat{x}_{t-1} + R(\tilde{\theta})z_t, \tag{6}$$

[1] $\widehat{x}_t$ denotes the percentage deviation of the generic variable x_t from a deterministic steady state x chosen as approximation point.

where z is a vector collecting the structural shocks of the DSGE model, while $T(\tilde{\theta})$ and $R(\tilde{\theta})$ are matrices one obtains after solving a DSGE model with standard solution techniques.

The impulse response functions of the variables in x to a structural shock i at horizon k φ_{ik}^D are given by:

$$\begin{aligned} \varphi_{i,0}^D &= R(\tilde{\theta})z_i, k = 0 & (7) \\ \varphi_{i,k}^D &= T(\tilde{\theta})\varphi_{k-1,i}^D, k = 1, 2, ...K. & (8) \end{aligned}$$

The vector of structural parameters of the DSGE model defined as in (6) does not contain any variances or covariances of a measurement error or any error term emerging from confronting the DSGE model with the data, but only the variances of the structural shocks. When the DSGE model is estimated by matching the corresponding impulse response functions, an additional error term occurs. Its variance covariance matrix is denoted by Ω and is also estimated. The vector comprising the vector of deep parameters $\tilde{\theta}$ and the vectorized Ω is defined as $\theta = [\ \tilde{\theta} \quad vec(\Omega)\ ]'$.

3.3 The idea in a nutshell

On the one hand, the distribution of the parameters of the DSGE model is estimated by matching the corresponding impulse response function of the VMA model. On the other hand, the distribution of structural impulse response functions of the VMA model are identified by applying sign restrictions which are derived from the DSGE model. Both distributions are therefore conditional distributions: they depend on a realization of the impulse response function of the VMA model and on restrictions from the DSGE model, i.e. a realization of a vector of structural parameters of the DSGE model, respectively. This section sets out how the conditional distributions can be combined to derive the joint distribution.

The joint posterior distribution of θ and φ, given a matrix with time series observations Y, $p(\theta, \varphi^V|Y)$, can be decomposed in different ways, depending on whether the DSGE model is employed to identify the VMA model or not. In the latter case the joint posterior is given by:

$$p(\varphi^V, \theta|Y) = p(\varphi^V|Y)p(\theta|\varphi^V). \quad (9)$$

This equation can be justified twofold: In the case that the DSGE model is estimated by matching the corresponding impulse response functions and not time series observations, the distribution of θ conditional on φ^V and Y is equal to the distribution of θ conditional on φ^V only[2]. The second justification is shown by Smith (1993) and DelNegro and

[2] It then holds:

$$p(\theta|\varphi^V, Y)p(\varphi^V|Y) = p(\varphi^V|Y)p(\theta|\varphi^V)$$

Schorfheide (2004) and discussed in section 3.4, when setting the framework in a broader context.

In the case that the likelihood of the VMA model impulse response functions depends on restrictions from the DSGE model, $p(\theta, \varphi^V|Y)$ is given as:

$$p(\varphi^V, \theta|Y) = p(\varphi^V|\theta, Y)p(\theta|Y). \tag{10}$$

The framework presented in this paper is based on the argument that both distributions are at least proportionally equal:

$$p(\varphi^V|Y)p(\theta|\varphi^V) \propto p(\varphi^V|\theta, Y)p(\theta|Y), \tag{11}$$

and can be approximated sufficiently well by Monte Carlo Markov Chain Methods.

Denote the Jacobian matrix collecting the derivatives of φ^V with respect to Φ by $J(\varphi^V \rightarrow A, \Phi)$. Considering the relationship between the coefficients of the VMA model and the impulse response function of the VMA model ($\Phi_i A = \varphi_i^V$):

$$p(\varphi^V|\theta, Y) = p(A, \Phi|\theta, Y)J(\varphi^V \rightarrow A, \Phi), \tag{12}$$

equation (11) is given by:

$$p(\varphi^V|Y)p(\theta|\varphi^V) \propto p(A, \Phi|\theta, Y)J(\varphi^V \rightarrow A, \Phi)p(\theta|Y). \tag{13}$$

Note that the conditional distributions of interest ($p(\theta|\varphi^V)$ and $p(A, \Phi|\theta, Y)J(\varphi^V \rightarrow A, \Phi)$ are on different sides of the proportionally sign in (13). It is therefore possible to employ a Gibbs sampling algorithm, i.e. to draw from two conditional distributions in order to evaluate the joint distribution. In the following section I will first relate the approach to existing methodologies before I describe both conditional distributions in detail.

3.4 Nested approaches

Taking a broader perspective, several closely-related methodologies evolve as special cases of this approach: the pure sign restriction approach of Uhlig (2005a), the DSGE-VAR methodology of DelNegro and Schorfheide (2004) and the case of probabilistic restrictions of Kociecki (2005).

The latter arises if the restrictions derived from the DSGE model are constant across the parameter space. Then it is possible to generate a prior distribution for the impulse response functions of the VAR model from the DSGE model and use it as a prior for the parameters of the VAR model. Since, as pointed out by Kociecki (2005), the sign restriction approach is a special case of the probabilistic approach, this methodology is also nested. The sign restriction approach arises if the prior distribution for some impulse response function exhibits a very small variance, i.e. determines the sign of this impulse. It is equivalent to using an indicator function placing zero probability weight on VAR model parameter regions whenever the a priori sign restrictions are not satisfied. Therefore, in the case that the DSGE model determines constant sign restrictions across the parameter space it is not necessary to draw from the conditional distribution of θ. One only needs to draw from $p(A, B|\theta, Y)$.

The DSGE-VAR methodology arises once the framework is rewritten in terms of the parameters instead of the impulse response functions of the VAR model, and in the case that the DSGE model is fully stochastically specified.

$$p(A, B|Y)p(\theta|A, B) \propto p(A, B|\theta, Y)p(\theta|Y) \tag{14}$$

The right-hand side is the expression used to evaluate the joint posterior distribution of $p(A, B, \theta|Y)$: since the DSGE model is fully stochastically specified it is possible to derive an analytical solution for the marginal posterior of θ. The decomposition on the left-hand side again legitimates the decomposition used in (9): the posterior distribution of the parameters of the VAR does not depend on the vector of structural parameters of the DSGE model. As argued in DelNegro and Schorfheide (2004) and Smith (1993), A and B can then be used to learn about the parameter vector θ.

4 Evaluating the joint distribution

In this section the conditional distributions employed in the estimation process are described in detail. I start by describing the distribution of the VMA model conditional on the parameter vector of the DSGE model. Then the distributions of the parameters of the DSGE conditional on the impulse response functions of the VMA model are set out.

4.1 Conditional distribution of the VMA model parameters

The conditional distribution described in this section is $p(\varphi^V|\theta, Y)$ from the right-hand side of (11). It is conditional since the prior distribution for the impulse response functions $p(\varphi^V) = p(\varphi^V|\theta)$ is derived from the DSGE model[3] The posterior distribution of the structural impulse responses φ^V is obtained by combining the coefficient estimates of the reduced-form VMA model Φ with an impulse matrix A. In order to write this distribution in terms of the reduced-form VMA model coefficients and the impulse matrix, it has to be scaled by the Jacobian $J(\varphi^V \to A, \Phi)$. The prior distribution for

[3]The impulse response functions of the DSGE model define a probability distribution of impulse response functions dependent on θ.

the structural impulse response function is set out and the Jacobian is derived in the first subsection.

Afterwards, the distribution $p(A, \Phi|\theta)$ is decomposed into a conditional distribution of the VMA model coefficients and a marginal distribution of the impulse matrix A:

$$p(\Phi, A|\theta) = p(\Phi|A, \theta)p(A|\theta). \tag{15}$$

Combining this prior distribution with the likelihood yields the posterior distribution $p(A, \Phi|\theta, Y)J(\varphi^V \to A, \Phi)$. The likelihood is described in the third part of this section. The resulting posterior distribution is difficult to evaluate for various reasons: given the restrictions, it is, to the best of my knowledge, not possible to draw the impulse matrix A of the VMA model for a reasonably large set of variables. It is not possible if only submatrices, i.e. fewer shocks than variables, are considered. This also implies that the distributions conditional on A are not defined, causing problems in the case that the restrictions are formulated for more than one period.

I therefore suggest in the fourth section deriving the restrictions from the DSGE model as sign restrictions. For each realization of the impulse response function of the DSGE model the corresponding sign restrictions are put on the VMA model. The coefficients of the VMA model are conditional on the impulse response functions of the DSGE model, similar to Uhlig (2005a), where the posterior distribution of the VAR parameters is multiplied by an indicator function that puts zero probability in parameter regions whenever the restrictions derived from the DSGE model are not satisfied. The distribution of parameters of the DSGE model θ defines a set of restrictions put on the parameters of the VMA model. This conditional prior distribution combined with the likelihood then yields the posterior distribution. A further simplification is considered in the concluding part of this section: the approximation of the VMA model by a VAR model.

4.1.1 The Jacobian $J(\varphi^V \to A, \Phi)$

Denote the impulse response functions in period k as φ_k^V. If all shocks are included, the matrix is of size $m \times m$, where the entry i, j corresponds to the response of variable i to an innovation in variable j. The prior for the impulse responses has to be specified for as many periods as there are impulse response functions to be estimated. The vectorized impulse responses are assumed to be normally distributed:

$$\begin{bmatrix} vec(\varphi_0) \\ vec(\varphi_1) \\ vec(\varphi_2) \\ \vdots \\ vec(\varphi_l) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} vec(\bar{\varphi}_0) \\ vec(\bar{\varphi}_1) \\ vec(\bar{\varphi}_2) \\ \vdots \\ vec(\bar{\varphi}_k) \end{bmatrix}, \begin{bmatrix} \bar{V}_{00} & \bar{V}_{01} & \bar{V}_{02} & \cdots & \bar{V}_{0l} \\ \bar{V}_{10} & \bar{V}_{11} & \bar{V}_{12} & \cdots & \bar{V}_{1k} \\ \bar{V}_{20} & \bar{V}_{21} & \bar{V}_{22} & \cdots & \bar{V}_{2k} \\ \vdots & & & & \\ \bar{V}_{k0} & \bar{V}_{k1} & \bar{V}_{k2} & \cdots & \bar{V}_{kk} \end{bmatrix} \right). \tag{16}$$

The probability distribution $p(\varphi_0, \varphi_1, ..., \varphi_k)$ can be decomposed into a marginal distribution of $p(\varphi_0)$ and succeeding conditional distributions:

$$p(\varphi_0, \varphi_1, ..., \varphi_k) = p(\varphi_k|\varphi_{k-1} \cdots \varphi_0)p(\varphi_{k-1}|\varphi_{k-2} \cdots \varphi_0) \cdots p(\varphi_1|\varphi_0)p(\varphi_0), \tag{17}$$

with

$$\begin{aligned} p(vec(\varphi_0)) &= \mathcal{N}(vec(\bar{\varphi}_0), \bar{\bar{V}}_{00}) & (18) \\ p(vec(\varphi_i|vec(\varphi_{i-1}) \cdots vec(\varphi_0)) &= \mathcal{N}(\theta_i, \Delta_{ii}), i = 1 \cdots k, & (19) \end{aligned}$$

and θ_i and Δ_{ii} abbreviate the usual definitions for conditional distributions:

$$\theta_i = vec(\bar{\varphi}) + \left[\ \bar{V}_{i0} \cdots \bar{V}_{ii-1}\ \right] \begin{bmatrix} \bar{V}_{00} & \cdots & \bar{V}_{0i-1} \\ \vdots & \ddots & \vdots \\ \bar{V}_{i-1,0} & \cdots & \bar{V}_{i-1,i-1} \end{bmatrix}^{-1} \begin{bmatrix} vec(\varphi_0 - \bar{\varphi}_0) \\ \vdots \\ vec(\varphi_{i-1} - \bar{\varphi}_{i-1}) \end{bmatrix}$$

$$\Delta_{ii} = \bar{V}_{ii} - \left[\ \bar{V}_{i0} \quad \cdots \quad \bar{V}_{ii-1}\ \right] \begin{bmatrix} \bar{V}_{00} & \cdots & \bar{V}_{0i-1} \\ \vdots & \ddots & \vdots \\ \bar{V}_{i-1,0} & \cdots & \bar{V}_{i-1,i-1} \end{bmatrix}^{-1} \begin{bmatrix} \bar{V}_{0i} \\ \vdots \\ \bar{V}_{i-1,i} \end{bmatrix}$$

In order to write the prior distribution in terms of the reduced form coefficients it is necessary to scale the probability distribution with the Jacobian:

$$p(\varphi) = p(f(\Phi))J(\varphi \Rightarrow \Phi). \tag{20}$$

The relationship between structural and reduced form moving average coefficients is given by:

$$\begin{aligned} \varphi_0 &= A & (21) \\ \varphi_i &= \Phi_i A, i = 1 \cdots k & (22) \end{aligned}$$

.

Note that I have left out Φ_0 since this matrix is normalized to an identity matrix by assumption. This also indicates that it is not possible to infer on φ_0 from the estimated reduced VMA model.

The Jacobian is calculated in the following way. Applying the vec-operator yields:[4]

$$vec(\varphi_i) = (A' \otimes I_{m \times m})vec(\Phi_i). \tag{23}$$

The Jacobian matrix is defined as:

$$J(\varphi \rightarrow \Phi) = det \begin{bmatrix} \frac{\partial vec(\varphi_1)}{\partial vec(\Phi_0)} & \frac{\partial vec(\varphi_1)}{\partial vec(\Phi_1)} & \cdots & \frac{\partial vec(\varphi_1)}{\partial vec(\Phi_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial vec(\varphi_k)}{\partial vec(\Phi_0)} & \frac{\partial vec(\varphi_k)}{\partial vec(\Phi_1)} & \cdots & \frac{\partial vec(\varphi_k)}{\partial vec(\Phi_k)} \end{bmatrix}. \tag{24}$$

[4]Note that $vec(AB) = (I \otimes A)vec(B) = (B' \otimes I)vec(A))$

Due to the fact that $\frac{\partial vec(\varphi_i)}{\partial vec(\Phi_j)} = 0$ for $j > i$, the matrix becomes a block triangular matrix and the determinant is given by:

$$
\begin{aligned}
J(\varphi \to \Phi) &= |\frac{\partial vec(\varphi_0)}{\partial vec(\Phi_0)}| \times |\frac{\partial vec(\varphi_1)}{\partial vec(\Phi_1)}| \cdots |\frac{\partial vec(\varphi_k)}{\partial vec(\Phi_k)}| \\
J(\varphi \to \Phi) &= |(A' \otimes I_{m\times m})|^k = |A|^{mk} \qquad (25)
\end{aligned}
$$

$$(26)$$

4.1.2 Decomposition of the distribution $p(\Phi, A) = p(\Phi|A)p(A)$

A prior distribution for the reduced form coefficients conditional on $\varphi_0 = A$ is formulated as:

$$p(A, \Phi_1, ..., \Phi_k) = p(\Phi_k|\Phi_{k-1} \cdots \Phi_0)p(\Phi_{k-1}|\Phi_{k-2} \cdots A) \cdots p(\varphi_1|A)p(A)J(\varphi \to \Phi), \quad (27)$$

where

$$
\begin{aligned}
p(vec(A)) &= \mathcal{N}(vec(\bar{\varphi}_0), \bar{\bar{V}}_{00}) & (28) \\
p(vec(\Phi_i)|vec(\Phi_{i-1}) \cdots vec(A)) &= \mathcal{N}(\bar{\Phi}_i, \bar{\bar{V}}_i i) & (29)
\end{aligned}
$$

with

$$
\begin{aligned}
\bar{\Phi}_i &= (A' \otimes I_{m\times m})\theta_i & (30) \\
\bar{\bar{V}}_i i &= (A^{-1'} \otimes I_{m\times m})\Delta_{ii}(A^{-1'} \otimes I_{m\times m}). & (31)
\end{aligned}
$$

4.1.3 The likelihood for the reduced-form coefficients

Consider the VMA(k) process:

$$Y_t = u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots + \Phi_k u_{t-k}. \quad (32)$$

This can be written in state space form:

$$
\begin{aligned}
\xi_{t+1} &= F\xi_t + U_{t+1} & (33) \\
y_t &= H\xi_t, & (34)
\end{aligned}
$$

where

$$\xi_t = \underset{m*k\times 1}{\begin{bmatrix} u_t & \cdots & u_{t-k} \end{bmatrix}}'$$

$$F = \underset{m*k\times m*k}{\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ I_m & 0 & \cdots & 0 & 0 \\ 0 & I_m & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_m & 0 \end{bmatrix}}$$

$$U_{t+1} = \underset{m*k\times 1}{\left[\begin{array}{cccc} u_{t+1} & 0 & \cdots & 0 \end{array}\right]'}$$

$$H = \underset{m*\times m*k}{\left[\begin{array}{cccc} I_m & \Phi_1 & \cdots & \Phi_k \end{array}\right]}.$$

Given an initial condition for y_0 and Σ_0, the likelihood can then be written as:

$$p(y_T, \ldots, y_0|\Phi_1, ..., \Phi_k, \Sigma) = p(y_T|y_{T-1} \ldots y_0, \Phi_1, ..., \Phi_k, \Sigma) \cdots p(y_0|\Phi_1, ..., \Phi_k, \Sigma), \quad (35)$$

where:

$$p(y_t|y_{t-1} \ldots y_0, \Phi_1, ..., \Phi_k) = \mathcal{N}(y_{t|t-1}, \Sigma_{t|t-1}). \quad (36)$$

$y_{t|t-1}$ and $\Sigma_{t|t-1}$ denote the optimal forecast at time t, which is a function of the coefficient matrices. The impulse matrix A is not part of the likelihood function, instead the variance covariance matrix $\Sigma = A'A$.

4.1.4 The posterior distribution

The posterior of the reduced form coefficients is derived by combining (35) and (27):

$$p(\Phi_1, \cdots \Phi_k, A|\theta, Y) = p(y_T, \ldots, y_0|\Phi_1, ..., \Phi_k, \Sigma)p(A, \Phi_1, ..., \Phi_k|\theta). \quad (37)$$

To identify the impulse matrix A from the likelihood estimate of the variance covariance matrix I utilize the prior distribution for $\varphi_0 = A$ derived from the DSGE model in the following way: the impulse matrix $\breve{A}$ is defined as a sub matrix of A of size $m \times n$ where n is the number of structural shocks under consideration, i.e. the structural shock of interest as well as other shocks necessary to distinguish this shock. These shocks have to be included in the DSGE model as well. In order to indicate that the restrictions put on A rely on the model and therefore its parameter vector θ, I write $\breve{A}(\theta)$. Given a number of rowvectors q_j forming an orthonormal matrix Q and the lower Cholesky decomposition of Σ, $\tilde{A}$, $\breve{A}(\theta)$ is defined as: $\breve{A}(\theta) = \tilde{A}Q(\theta)$.

Every realization of the vector of the parameters of the DSGE model θ is associated with an impulse response function of the DSGE model and a realization of $\breve{A}(\theta)$. A sequence of realizations of θ yields a sequence of restrictions and therefore a related prior probability distribution. Given a realization of an impulse response function of the DSGE model φ^D the posterior distribution is evaluated the following way:

1. Derive the sign restrictions from φ^D.
2. Draw a realization of Φ and Σ from the distribution (37).
3. Calculate $\tilde{A}$ and draw $Q(\theta)$ from a uniform distribution such that $\breve{A}(\theta) = \tilde{A}Q(\theta)$ fulfils the sign restriction.
4. Given A, compute the structural impulse responses from $\varphi_i = \Phi_i A, i = 1 \cdots k$.

4.1.5 The conditional distribution of the VAR model

Estimating a VAR model is less complicated. In practice whenever possible, i.e. if the VMA model is stationary, it is approximated by a VAR model[5]. In this section I therefore briefly lay out the approach for this case.

As shown by Uhlig (1997), the prior distribution for B and Σ can be specified choosing appropriate B_0, N_0, S_0, v_0 as:

$$\begin{aligned} vec(B)|\Sigma &\sim \mathcal{N}(vec(B_0), \Sigma \otimes N_0^{-1}) && (38) \\ \Sigma &\sim \mathcal{IW}(v_0 S_0, v_0). && (39) \end{aligned}$$

Denote the maximum likelihood estimates of Σ and B as $\tilde{\Sigma} = \frac{1}{T}(Y - X\hat{B})'(Y - X\hat{B})$ and $\hat{B} = (X'X)^{-1}X'Y$. The posterior is then given as[6]:

$$\begin{aligned} vec(B)|\Sigma &\sim \mathcal{N}(vec(B_T), \Sigma \otimes N_T^{-1}) && (40) \\ \Sigma &\sim \mathcal{IW}(v_T S_T, v_T), && (41) \end{aligned}$$

where

$$\begin{aligned} N_T &= N_0 + X'X && (42) \\ B_T &= N_T^{-1}(N_0 B_0 + X'X\hat{B}) && (43) \\ S_T &= \frac{v_0}{v_T} S_0 + \frac{T}{v_T}\tilde{\Sigma} - \frac{1}{v_T}(B_0 - \hat{B})' N_0 N_T^{-1} X'X (B_0 - \hat{B}) && (44) \\ v_T &= v_0 + T. && (45) \end{aligned}$$

Drawing from a joint posterior of B, Σ and $\breve{A}(\theta)$ is conducted in the following steps:

1. The impulse responses of the DSGE determine the restrictions put on $\breve{A}(\theta)$.
2. Draw B and Σ from the posterior (40) and (41).
3. Calculate $\tilde{A}$ and draw $Q(\theta)$ from a uniform distribution such that $\breve{A}(\theta) = \tilde{A}Q(\theta)$ fulfills the sign restriction.

4.2 The conditional distribution of the DSGE model parameters

Since the DSGE model is not assumed to be a proper representation of the data-generating process, the structural parameters are not estimated by matching the data

[5] I will employ the expression VAR model in the remaining sections too.

[6] A formal derivation is given in appendix A

Y. Instead, the DSGE model is assumed to replicate the implied dynamics of the data, i.e. the impulse response functions of the VAR model. This induces matching a given realization of the impulse response function of the VAR model to the i-th shock at horizon k, $\varphi^V_{i,k}$:

$$\varphi^V_{i,k} = \varphi^D_{i,k}(\tilde{\theta}) + \omega_{i,k}. \tag{46}$$

Stacking the impulse response functions over $1, .., K$ periods together yields:

$$\varphi^V_i = \varphi^D_i(\tilde{\theta}) + \omega_i \tag{47}$$

with all vectors of dimension $m{*}k{\times}1$. The error term ω_i has the property $E[\omega_i\omega_i'] = \Omega_{\omega_i}$, which is part of the vector θ. Since the structural shocks are assumed to be independent, the probability of $p(\theta|\varphi^V)$ can be written as:

$$p(\theta|\varphi^V) = p(\theta|\varphi^V_1, \varphi^V_2, \cdots \varphi^V_i) = p(\theta|\varphi^V_1)p(\theta|\varphi^V_2)\cdots p(\theta|\varphi^V_i). \tag{48}$$

The vector θ is estimated in two steps: First Ω_{ω_i} is estimated, and afterwards the vector of deep parameters $\tilde{\theta}$. The variance covariance is estimated by making use of the relationship:

$$\omega_i = \varphi^V_i - \varphi^D_i(\tilde{\theta}). \tag{49}$$

For every realization of φ^V_i a reasonable number of draws from $p(\theta)$ is taken[7], and the corresponding impulse response function $\varphi^D_i(\tilde{\theta})$ and the error terms are computed. $\tilde{\Omega}_{\omega_i}$ is then estimated as the covariance matrix of these error terms. For each shock i the likelihood $l_i\left(\tilde{\theta}|\varphi^V_i, \tilde{\Omega}_{\omega_i}\right)$ is given by:

$$l_i\left(\tilde{\theta}|\varphi^V_i, \tilde{\Omega}_{\omega_i}\right) = -\frac{Km}{2}ln(2\pi) - \frac{1}{2}ln(|\tilde{\Omega}_{\omega_i} \otimes I_K|) - \frac{1}{2}(\omega_i)'\left(\tilde{\Omega}_{\omega_i} \otimes I_K\right)^{-1}(\omega_i). \tag{50}$$

Combining this likelihood with a prior distribution for θ yields the posterior distribution.

One potential issue arising when matching impulse response functions of a DSGE model and a VAR model was pointed out by McGrattan, Chari, and Kehoe (2005): the implied VAR model representation of the DSGE model might be of infinite order but the empirical VAR model is often of lower order. One solution, suggested by Cogley and Nason (1995), is to simulate artificial time series from the DSGE model, estimate a VAR model from the artificial time series and compare this VAR model to the VAR model estimated from the actual data.

4.3 Sampling algorithm for the joint posterior distribution

In order to evaluate the joint posterior distribution of the parameters of the DSGE model and the VAR model I propose a Gibbs sampling algorithm combined with a

[7] In the simulation and estimation I used 50 draws per realization.

Metropolis-Hastings step. The Gibbs sampling algorithm allows to draw from the conditional distributions laid out in detail in sections 4.1.4 and 4.2. The Metropolis-Hastings step is an acceptance/rejection sampling algorithm that determines the probability space in which the implied impulse response functions of the DSGE model and those of the VAR model coincide. It is carried out I times.

The algorithm can roughly be summarized in the following way. At each iteration $i = 1, \ldots, I$ d-draws are taken from the conditional densities $p(\theta|\varphi^V)^i$ and $p(\varphi^V|\theta, Y)^i$.[8] These draws form candidate distributions $p(\theta|\varphi^V)^{\tilde{i}}$ and $p(\varphi^V|Y, \theta)^{\tilde{i}}$. Via acceptance/rejection, the candidate distributions are compared with $p(\theta|\varphi^V)^i$ and $p(\varphi^V|\theta, Y)^i$. Draws with higher posterior density are kept and form the new densities $p(\theta|\varphi^V)^{i+1}$ and $p(\varphi^V|\theta, Y)^{i+1}$. More precisely, at each iteration $i = 1, \ldots, I$ the algorithm involves the following steps:

1. Draw $j = 1 \ldots d$ times from $p(\theta|\varphi^V)^i$.
2. For every realization θ_j of the vector of deep parameters of the DSGE model derive the corresponding sign restriction.
3. Draw Σ_j from (41) and B_j from (40). Compute the lower Cholesky decomposition and find an $\breve{A}_j = \tilde{A}_j Q_j$ fulfilling the sign restrictions from $\varphi_j^D(\tilde{\theta}_j)$. Compute φ_j^V, yielding $p(\varphi^V|Y, \theta)^{\tilde{i}}$.
4. For every realization of φ_j^V derived in step 3 find the θ that maximizes (50) combined with the prior $p(\theta)$. This yields $p(\theta|\varphi^V)^{\tilde{i}}$.
5. Do acceptance-rejection by comparing $p(\theta|\varphi^V)^{\tilde{i}}$ with $p(\theta|\varphi^V)^{i-1}$. Keep the corresponding vectors from $p(\varphi^V|\theta)^{\tilde{i}}$. This yields $p(\theta|\varphi^V)^{i+1}$ and $p(\varphi^V|\theta)^{i+1}$.
6. Start again at 1.

The chain converges if $p(\theta|\varphi^V)^i$ and $p(\theta|\varphi^V)^{i-1}$ and also $p(\varphi^V|\theta)^i$ and $p(\varphi^V|\theta)^{i-1}$ are similar, i.e. the acceptance rate is low.

In the remaining sections of the paper I will discuss the properties of the sampling algorithm in more detail using a Monte Carlo experiment, i.e. specify precisely the number of iterations and the convergence of the algorithm. Afterwards I will put the methodology to work and confront it with the data.

5 Example 1: A Monte Carlo Experiment

In order to illustrate the methodology suggested above I use a simple fiscal theory of the price level (FTPL) model as described in Leeper (1991) to identify the response of

[8] In the first iteration step $p(\theta|\varphi^V)^1 = p(\theta)$.

inflation to a monetary policy shock, i.e. an unexpected increase in the interest rate. The FTPL model is chosen because it can be reduced to two equations in real debt and inflation. It is the most simple DSGE model exhibiting different signs of the impulse response functions depending on two parameters only. Furthermore, the solution and properties of the FTPL model are well known by economists, which makes the example very transparent.

I simulate data from the FTPL model and using the methodology outlined above show that the 'true' signs of the impulse response functions and the corresponding distribution of the parameters of the FTPL model are found, even if the chain is initialized with a wrong guess.

5.1 The FTPL model

The representative household maximizes its utility in consumption [9] c_F and real money balances m_F:

$$U_t = log(c_{F,t}) + log(m_{F,t}) \tag{51}$$

subject to the budget constraint:

$$c_{F,t} + m_{F,t} + b_{F,t} + \tau_{F,t} = yF + \frac{1}{\pi_{F,t}} m_{F,t-1} + \frac{R_{F,t-1}}{\pi_{F,t}} b_{F,t}, \tag{52}$$

where b_F denotes bond holdings, τ_F lump sum taxes, y_F income, R_F nominal interest rates and π_F inflation. Small letters denote real variables, capital letters nominal variables.

The government has to finance government expenditures g_F by issuing bonds, collecting taxes and seignorage. The budget constraint is therefore given by:

$$b_{F,t} + m_{F,t} + \tau_{F,t} = g_F + \frac{M_{F,t-1}}{P_{F,t}} + R_{F,t-1} \frac{B_{F,t-1}}{P_{F,t}}. \tag{53}$$

The monetary authority sets the nominal interest rate R_F following the interest rate rule:

$$R_{F,t} = \alpha_{F0} + \alpha_F \pi_{F,t} + z_{F,t}, \tag{54}$$

where α_{0F} and α_F are policy coefficients. z_F denotes a monetary policy shock, specified as

$$z_{F,t} = \rho_{F,1} z_{F,t-1} + \epsilon_{F1,t} \tag{55}$$

$$\epsilon_{F1,t} \sim N(0, \sigma_{F1}). \tag{56}$$

The fiscal authority sets taxes according to:

$$\tau_{F,t} = \gamma_{F0} + \gamma_F b_{F,t-1} + \psi_{F,t}, \tag{57}$$

[9] All variables and parameters associated with the FTPL model are labeled with a F.

where again γ_{F0} and γ denote policy coefficients. The innovation in fiscal policy has the following characteristics:

$$\psi_{F,t} = \rho_F 2\psi_{F,t-1} + \epsilon_{F2,t} \quad (58)$$
$$\epsilon_{F2,t} \sim N(0, \sigma_{F2}). \quad (59)$$

The model can be linearized and summarized by two equations[10]:

$$\tilde{\pi}_{F,t+1} = \beta_F \alpha_F \tilde{\pi}_{F,t} + \beta_F z_{F,t} \quad (60)$$
$$\tilde{b}_{F,t} + \varphi_{F1}\tilde{\pi}_{F,t} + \varphi_{F3} z_{F,t} + \psi_{F,t} = (\beta_F^{-1} - \gamma_F)\tilde{b}_{F,t-1} - \varphi_{F4} z_{F,t-1} - \varphi_{F2}\tilde{\pi}_{F,t-1}. \quad (61)$$

5.2 Dynamics of the FTPL model

The dynamics of the system depend on whether fiscal and monetary policy are active or passive, i.e. they depend on the policy parameters α_F and γ_F only. Different policy regimes emerge for:

- $|\alpha_F \beta_F| > 1$ and $|\beta_F^{-1} - \gamma_F| < 1$ for active monetary (AM) and passive fiscal (PF) policy. This will be referred to as regime I.
- $|\alpha_F \beta_F| < 1$ and $|\beta_F^{-1} - \gamma_F| > 1$ for active fiscal (AF) and passive monetary (PM) policy. This will be referred to as regime II.
- AM/AF and PF/PM. These cases are not considered here.

Both policy regimes imply different signs of the impulse response function for inflation and real debt. In regime I a monetary policy shock (an unanticipated increase in the nominal interest rate) will lead to a negative response of inflation and a positive response of real debt. A fiscal policy shock (an unanticipated increase in taxes) will have no effect on inflation and decrease the real debt. In regime II, a monetary policy shock leads to an increase in inflation and an initial decrease in real debt. A fiscal policy shock has a negative effect on both variables. Impulse response functions for each shock, regime and variable are plotted in appendix E together with the corresponding distributions of α_F and γ_F .

5.3 Specification and Identification of the VAR

The VAR model consists of two variables, inflation π_F and real debt b_F, with no constant or time trend: $y_{F,t} = [\pi_{F,t} b_{F,t}]'$. The VAR model with one lag is given by:

$$y_{F,t} = B y_{F,t-1} + u_{F,t}$$
$$E[u_{F,t} u'_{F,t}] = \Sigma_F.$$

[10] See appendix B for a derivation.

Ordering the fiscal policy shock first and the monetary policy shock second, based on the model the following characteristics of the impulse matrix A_F have to hold:

- If regime I holds:
  - Fiscal policy shock: $A_{F,11} = 0$ $A_{F,21} < 0$.
  - Monetary policy shock: $A_{F,21} < 0$ and $A_{F,22} > 0$.
- If regime II holds:
  - Fiscal policy shock: $A_{F,11} < 0$ $A_{F,21} < 0$.
  - Monetary policy shock: $A_{F,21} > 0$ and $A_{F,22} < 0$.

Since the sign of the reaction of real debt to a monetary policy shock does not identify the shock in the case of regime II, the monetary policy shock is ordered second, implying that both variables have to fulfil the sign restriction for a fiscal policy shock first. Then the sign of the response of real debt is restricted, while the response of inflation is left open.

5.4 A Monte Carlo Experiment

I simulate data from the model over 200 periods with $\alpha_F = 0.5$ and$\gamma_F = -0.00001$, i.e. the case of active fiscal and passive monetary policy. I choose the prior distribution of α_F and γ_F based on estimates by Davig and Leeper (2005):

Parameter	mean(I)	standard deviation(I)	mean(II)	standard deviation(II)
α_F	1.308	0.253	0.522	0.175
γ_F	0.0136	0.012	-0.0094	0.013

Table 1: Prior distribution for parameters of the FTPL model

The prior distribution is plotted in figure 1. The model fulfills the requirements for investigating the question of how inflation reacts after a monetary policy shock: depending on the parameterization it allows for qualitatively different reactions of inflation, and the DSGE model incorporates all other shocks necessary, here the fiscal policy shock, to distinguish the shock of interest. The corresponding impulse responses for each regime are plotted in appendix E: Figures 2 and 3 provide Bayesian impulse response plots for draws from the prior distribution of regime I and figures 4 and 5 for draws from the prior distribution of regime II.

The sampling algorithm is specified by setting $d = 200$ to approximate the candidate distribution. Afterwards, $I = 50$ and 200 draws are taken at each iteration. Since the data are simulated from regime II, the outcome expected is the distribution of regime

II, with the corresponding impulse responses of inflation and real debt for a fiscal policy shock and real debt for a monetary policy shock. Furthermore, inflation should rise in response to a monetary policy shock.

As figure 6 indicates, this is indeed the case, even though I initialize the chain with a wrong guess. The posterior distribution of α_F and γ_F stems from regime II only. Figure 7 shows the response to a fiscal policy shock and figure 8 the response to a monetary policy shock. Inflation is indeed increasing.

6 Example 2: Application to the data

In this section I take the methodology to the data. Since the FTPL model is too stylized I introduce another very simple DSGE model: the deep habit model. This model was employed by Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008) to answer the question whether a simple model can account for the so-called price puzzle, i.e. the increase of inflation after a contractionary monetary policy shock. This DSGE model is able to generate a positive as well as a negative response of inflation after a monetary policy shock.

I flip the question and explore whether prices increase or decrease after a monetary policy shock using the methodology set out in this paper. I identify the monetary policy shock by employing sign restrictions from the DSGE model. The response of inflation will be left open.

In the remaining part of the section, first the DSGE model is set up and its dynamics are described. Finally, the DSGE model and the VAR model are estimated jointly.[11]

6.1 Deep habits model

The DSGE model consists of households, firms and a monetary authority. In the following, these parts of the DSGE model are characterized, the equilibrium is defined and the impulse response functions are analyzed. All variables and parameters associated with the deep habits model are labeled with an H.

[11] The (uncommented) matlab codes are available upon request and will (hopefully) be made available commented on my webpage soon.

6.1.1 Households

There is a continuum of households indexed by $j \in [0,1]$, which are all identical and infinitively lived. Household j's preferences are given by:

$$U_0^j = E_0 \sum_{t=0}^{\infty} \beta_H^t \left[\frac{1}{1-\sigma_H} x_{H,t}^j - \frac{\gamma_H}{1+\kappa_H} (h_{H,t}^j)^{1+\kappa_H} \right] \tag{62}$$

$$x_{H,t}^j = \left[\int_0^1 (c_{Hi,t}^j - \theta_H^d c_{Hi,t-1})^{1-\frac{1}{\eta_H}} di \right]^{\frac{1}{1-1/\eta_H}} \tag{63}$$

$$c_{Hi,t} = \int_0^1 c_{Hi,t}^j dj \tag{64}$$

where β_H denotes the discount factor, κ_H the inverse of the Frisch elasticity of labor supply, γ_H a preference weight on households j's labor supply and $x_{H,t}^j$ denotes the consumption basket. Equation (63) defines the deep habit: consumption of variety i is related to the past aggregate of this variety. Deep habits therefore imply that the level of marginal utility of individual goods varies. The aggregate is assumed to be exogenously given. The parameter θ_H^d measures the importance of the habit.

Demand for c_{it}^j is given by:

$$c_{Hi,t}^j = \left(\frac{P_{Hi,t}}{P_{H,t}} x_{H,t}^j + \theta_H^d c_{Hi,t-1} \right), \tag{65}$$

where $P_{H,t}$ denotes an aggregate price index:

$$P_{H,t} = \left[\int_0^1 P_{Hi,t}^{1-\eta_H} di \right]^{1/(1-\eta_H)}. \tag{66}$$

Households act as monopolistically competitive labor unions in the labor market earning the wage rate W_H^j. They face costs of changing wages ζ_{Hw}, which are quadratic in the deviation of nominal wage growth from an index factor $\tilde{\pi}_{Hw,t}$. The index factor evolves according to:

$$\tilde{\pi}_{Hw,t} = \vartheta_{Hw} \pi_{Hw}^{\star} + (1-\nu_{Hw}) \pi_{Hw,t-1}, \tag{67}$$

where ν_{Hw} measures the extent of wage indexation.

Households own firms and receive dividends $D_{H,t}^j$, and furthermore have access to a nominal risk-free bond B_H yielding the gross nominal interest rate R_H.

They maximize utility with respect to the budget constraint:

$$P_{H,t} x_{H,t}^j + \theta_H^d \int_0^1 P_{Hi,t} c_{Hi,t-1} di + B_{H,t}^j \tag{68}$$

$$= R_{H,t-1} B_{H,t-1}^j + W_{H,t}^j h_{H,t}^j + D_{H,t}^j - P_{H,t} \frac{\zeta_{Hw}}{2} \left(\frac{W_{H,t}^j}{W_{H,t-1}^j} - \tilde{\pi}_{Hw,t} \right).$$

6.1.2 Firms

Firms are monopolistically competitive. Firm i produces output using the following production function:

$$y_{Hi,t} = h_{Hi,t}. \tag{69}$$

Labor input is defined as:

$$h_{Hi,t} = \left(\int_0^1 (h^j_{Hi,t})^{1-1/\psi_H} dj \right)^{1/(1-1/\phi_H)}. \tag{70}$$

Given the price $W^j_{H,t}$ for $h^j_{H,t}$, the labor demand function is given by:

$$h^j_{Hi,t} = \left(\frac{W^j_{H,t}}{W_{H,t}} \right)^{-\psi_H} h_{Hi,t}, \tag{71}$$

where the aggregate wage rate $W_{H,t}$ is defined as:

$$W_{H,t} = \left[\int_0^1 W^{1-\psi_H}_{Hj,t} dj \right]^{1/(1-\psi_H)}. \tag{72}$$

Aggregating (71) yields the demand for household j's labor:

$$h^j_{H,t} = \left(\frac{W^j_{H,t}}{W_{H,t}} \right)^{-\psi_H} h_{H,t}. \tag{73}$$

Aggregating across consumers, the demand function for firm i's product is given by:

$$\begin{aligned} c_{Hi,t} &= \left(\frac{P_{Hi,t}}{P_{H,t}} \right)^{-\eta_H} x_{H,t} + \theta^d_H \theta^d_H c_{Hi,t-1} \\ c_{Hi,t} &= \int_0^1 c^j_{Hi,t} dj \\ x_{H,t} &= \int_0^1 x^j_{H,t} dj. \end{aligned} \tag{74}$$

From (74) the main mechanism becomes apparent: firms have an incentive to lower prices today if they expect future demand to be high relative to current demand. Additionally, the firm increases its weight on the price elastic term and therefore its price elasticity of demand.

Firms face quadratic adjustment costs ζ_{Hp} when changing nominal prices. Firms' profits are discounted by the following discount factor:

$$q_{H,t} = \beta^t_H \frac{x^{-\sigma_H}_{H,t}}{P_{H,t}}.$$

The maximization problem of firm i thus reads:

$$\underset{P_{Hi,t}}{m} ax E_0 = \sum_{t=0}^{\infty} q_{H,t} D_{Hi,t} \tag{75}$$

$$D_{Hi,t} = P_{Hi,t} c_{Hi,t} - W_{H,t} h_{Hi,t} - \frac{\zeta_{Hp}}{2} P_{H,t} \left(\frac{P_{Hi,t}}{P_{Hi,t-1}} - \tilde{\pi}_{H,t} \right)^2 .$$

Nominal prices are indexed by $\tilde{\pi}_{H,t}$, which evolves according to:

$$\tilde{\pi}_{H,t} = \nu_{Hp} \pi_H^\star + (1 - \nu_{Hp}) \pi_{H,t-1}. \tag{76}$$

6.1.3 Monetary Policy and market clearing

Monetary policy aims at stabilizing deviations in inflation and output from their steady state values $\pi_H^\star$ and $y_H^\star$. It sets the policy coefficients ρ_{Hr}, $\alpha_{H,\pi}$ and $\alpha_{H,y}$ according to the simple interest rate rule:

$$R_{H,t} = R_H^\star + \rho_{H,r}(R_{H,t-1} - R_H^\star) + (1 - \rho_{Hr}) \left[\alpha_{H,\pi}(\pi_{H,t} - \pi_H^\star) + \alpha_{Hy} \left(\frac{y_{H,t} - y_H^\star}{y_H^\star} + \epsilon_{H,t}. \right) \right] . \tag{77}$$

$\epsilon_{H,t}$ denotes the monetary policy shock: $\epsilon_{H,t} \sim \mathcal{N}(0, \sigma_{HR})$.

Market clearing implies:

$$h_{H,t}^j = \int_0^1 h_{Hi,t}^j di \tag{78}$$

$$h_{Hi,t} = \int_0^1 h_{Hi,t}^j dj \tag{79}$$

as well as:

$$c_{H,t} = y_{H,t}. \tag{80}$$

The aggregate resource constraint is given by:

$$y_{H,t} = h_{H,t}. \tag{81}$$

6.1.4 Equilibrium definition

I follow Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008) by concentrating on the symmetric equilibrium in which all consumers make the same choice over consumption, set the same wage and all firms set the same prices.

A recursive equilibrium is then defined as follows:

Definition 1 *Given initial values* $P_{H,0} > 0$ *and* $W_{H,0}$*, the recursive laws of motion for price and wage indexation* (76) *and* (67) *and a monetary policy, a rational expectations equilibrium (REE) for* $R_{H,t} \geq 1$*, is a set of sequences* $\{y_{H,t},\ c_{H,t},\ h_{H,t},\ x_{H,t},\ w_{H,t},\ P_{H,t}, R_{H,t}\}_{t=t_0}^{\infty}$

(i) that solve the firms' problem (75) *with s.t.* (74)*,*

(ii) that maximize households' utility (62) *s.t.* (73)*,* (68) *and a No-Ponzi-scheme condition,*

(iii) that clear the goods market (80) *and labor market, i.e.* (79) *and* (78) *hold,*

(iv) and that satisfy the aggregate resource constraint (81)*.*

The DSGE model is loglinearized around its steady state. An overview of the steady state and the loglinearized equations are given in Appendix C.1 and C.2 respectively.

6.1.5 Prior distribution of the parameters and impulse response functions

I estimate only those structural parameters crucial for the response of inflation[12]. For those parameters, prior distributions are specified which allow for a wide range of impulse response functions of the deep habits model. The parameters not estimated are calibrated as in Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008). An overview can be found in appendix C.3.

6.2 Estimation

In figure 9 the impulse response functions of the DSGE model when drawing from the prior distribution are plotted.[13] The signs of all the impulse response functions except the response of interest (inflation) are constant, i.e. for every draw from the parameter distribution of the DSGE model consumption, real wages and output will be decreasing while the interest rate increases. In order to distinguish the characterization of the shock from other shocks, I compare the combination of signs with combinations implied by other common shocks. These shocks are taken from Smets and Wouters (2003). The sign restriction of the monetary policy shock implied by the deep habits model are different from the signs of common shocks except for the price markup shock in Smets and Wouters (2003). Even though it is the shock exhibiting the smallest variance, I further include adjusted reserves as well as the price index of crude materials into the VAR model to distinguish the estimated shock (following Mountford and Uhlig

[12]Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008) also only estimate a subset of the structural parameters.

[13]All figures are provided in appendix E.

(2005)). While the former is restricted to react negativly, the latter is left unrestricted. Since both variables have no counterparts in the DSGE model, they are not matched. Overall, the VAR model consists of 7 variables: real GDP, real personal consumption, real wages, interest rates, adjusted reserves, the GDP deflator and the price index of crude materials. A complete description of the time series is given in Appendix D. The prior distribution of the VAR model is specified as a flat prior.

Before the DSGE model is estimated, I perform a Monte Carlo experiment to ensure the validity of the methodology, the identification and the specification of the sampling algorithm. The candidate distribution for the vector of deep parameters will be the prior distribution. In the Monte Carlo experiment I set $I = 20$ and draw $n = 200$ times at each iteration. First only a subvector of the parameters of the DSGE model consisting of θ_d, η, ζ_w, and ζ_p is estimated. The results are displayed in table 2 (columns 6 and 7) of appendix C.3 and show that all parameters are estimated very precisely around their true values (column 5). This is a very encouraging result, especially since the prior distribution is not centered around the true value.

Adding more parameters to the vector of estimated parameters has two effects. This is demonstrated by supplementing the vector of structural parameters with the coefficients of the Taylor rule $(\rho_{Hr}, \alpha_{H\pi}, \alpha_{Hy})$ and the inflation indexation parameter ν_{Hp}. On the one hand, this increases the flexibility of the DSGE model and therefore increases the ability to fit the impulse response functions of the data. Figure 10 provides plots of the impulse response function of the DSGE model and the VAR model. Both coincide and, more importantly, the 'true' impulse response function for inflation, i.e. the impulse response function for the parameter vector at which the DSGE model is simulated, is estimated. On the other hand, as shown in table 2 columns 8 and 9, the precision of the estimation is slightly blurred.

Given the encouraging results, I take the methodology to the data. At every iteration I take $n = 200$ draws, the number of iterations is set to $I = 20$. Table 2 column 10 and 11 report the mean and the standard deviation of the posterior distribution respectively. The estimation results for the posterior mean of some of the parameters of the DSGE model are very similar to those obtained by Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008)[14]: $\eta_H = 2.47$ (2.48), $\zeta_{Hp} = 14.89$ (14.47), $\zeta_{Hp} = 42.50$ (40.89),$\alpha_{Hr} = 0.01(0.04)$. I find slightly different estimates for the deep habit parameter $\theta_H^d = 0.72$ (0.85), the inflation indexation parameter $\nu_{Hp} = 0.1$ (0), and the policy coefficients $\rho_{Hr} = 0.81$ (0.74) and $\alpha_{H\pi} = 1.56, (1.26)$. Figure 12 displays the impulse response functions: since the parameters of the DSGE model are estimated similarly, the response of inflation is positive and significant for 68% probability bands.[15] However, while the graph indicates a positive response for the mean response of the VAR model, the uncertainty bands give rise to the conclusion that a negative response of inflation to a monetary shock is as likely as positive one.

[14]For comparison I report their findings in brackets after my estimates.

[15]It is not significant for 100% probability bands.

7 Conclusion

This paper has laid out a methodology for identifying the structural shocks of a Vector Autoregression (VAR) model while at the same time estimating a Dynamic Stochastic General Equilibrium (DSGE) model that is not assumed to replicate the data-generating process. To this end it has presented a framework for jointly estimating the parameters of a VAR model and a DSGE model.

The VAR model is identified based on restrictions from the DSGE model, i.e identification relies on restrictions explicitly derived from theory. This ensures consistency of the identification of the VAR model with the implied structural form of the DSGE model. Restrictions are formulated as sign restrictions. Thus, the DSGE model serves as a way to summarize the ideas economists have about the economy. Ideally, it incorporates the assumptions the researcher wants to discriminate between, but in any case it should be as agnostic as possible about the response of the variables of interest to the shock of interest.

The DSGE model is estimated by matching the impulse response functions of the VAR and of the DSGE, i.e. their implied dynamics. Therefore, it need not be a representation of the data-generating process. While the shock of interest has to be included, as well as other shocks necessary to distinguish it, the DSGE model need not be fully stochastically specified.

The methodology has been first illustrated by means of a Monte Carlo experiment and has been applied to the data afterwards. In the Monte Carlo experiment, artificial data has been simulated from a simple fiscal theory of the price level model in which fiscal policy is active and monetary policy passive. The sign of the response of inflation to a monetary policy shock has been investigated. Depending on the policy regime, i.e. the reaction coefficients of the policy rules, the response can either be negative or positive. The prior distributions of the policy parameters have been chosen such as to ensure that both regimes and therefore both responses are equally likely. The estimated impulse response function of the VAR model as well as the posterior distribution of the parameter of the DSGE model indicate that the methodology works correctly: the response of inflation shows the 'true' sign and the posterior distribution of the parameter of the DSGE model consists solely of policy coefficients from active fiscal and passive monetary policy.

Finally, the methodology has been used to estimate the response of inflation to a monetary policy shock. As a DSGE model, the deep habits model laid out by Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008) has been employed. The posterior estimates of the parameters of the DSGE model are similar or only slightly different from those obtained by the authors. Correspondingly, I find a positive and on a 68% level significant response of inflation to a monetary policy shock. However, while the mean of the impulse response function of the VAR model is positive, the uncertainty bands indicate that a negative response of inflation to a monetary policy is as likely as a positive one.

References

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2002): "Technology Shocks and Aggregate Fluctuations," draft, Northwestern University.

Beyer, A., and R. E. Farmer (2006): "Identification Problems in SDGE Models with an illustration to a small Macro model," Computing in Economics and Finance 2006 81, Society for Computational Economics.

Canova, F., and L. Sala (2006): "Back to Square One: Identification Issues in DSGE Models," Working Papers 303, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary policy shocks: What have we learned and to what end?," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1 of *Handbook of Macroeconomics*, chap. 2, pp. 65–148. Elsevier.

——— (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Cogley, T., and J. M. Nason (1995): "Output Dynamics in Real-Business-Cycle Models," *American Economic Review*, 85(3), 492–511.

Davig, T., and E. M. Leeper (2005): "Fluctuating Macro Policies and the Fiscal Theory," NBER Working Papers 11212, National Bureau of Economic Research, Inc.

DelNegro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARS," *International Economic Review*, 45(2), 643–673.

Faust, J. (1998): "The robustness of identified VAR conclusions about money," International Finance Discussion Papers 610, Board of Governors of the Federal Reserve System (U.S.).

Geweke, J. (1999): "Computational experiments and reality," Discussion paper, University of Minnesota and Federal Reserve Bank of Minneapolis.

Ingram, B. F., and C. H. Whiteman (1994): "Supplanting the 'Minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors," *Journal of Monetary Economics*, 34(3), 497–510.

Kociecki, A. (2005): "Priors For Impulse Responses in Bayesian Structural VAR Models," Discussion paper, National Bank of Poland.

Lanne, M., and H. Lütkepohl (2005): "Structural Vector Autoregressions with Nonnormal Residuals," Economics Working Papers ECO2005/25, European University Institute.

——— (2008): "Identifying Monetary Policy Shocks via Changes in Volatility," *Journal of Money, Credit and Banking*, 40(6), 1131–1149.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2009): "Structural Vector Autoregressions with Markov Switching," Economics Working Papers ECO2009/06, European University Institute.

Leamer, E. (1978): *Specification Searches*. John Wiley and Sons.

Leeper, E. M. (1991): "Equilibria under 'active' and 'passive' monetary and fiscal policies," *Journal of Monetary Economics*, 27(1), 129–147.

McGrattan, E., V. V. Chari, and P. Kehoe (2005): "Are Structural VARs Useful Guides for Developing Business Cycle Theories?," 2005 Meeting Papers 664, Society for Economic Dynamics.

Mertens, K., and M. O. Ravn (2008): "The Aggregate Effects of Anticipated and Unanticipated U.S. Tax Policy Shocks: Theory and Empirical Evidence," Economics Working Papers ECO2008/05, European University Institute.

Mountford, A., and H. Uhlig (2005): "What are the Effects of Fiscal Policy Shocks?," SFB 649 Discussion Papers SFB649DP2005-039, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Ravn, M. O., S. Schmitt-Grohé, and M. Uribe (2007): "Explaining the Effects of Government Spending Shocks on Consumption and the Real Exchange Rate," NBER Working Papers 13328, National Bureau of Economic Research, Inc.

Ravn, M. O., S. Schmitt-Grohé, M. Uribe, and L. Uuskula (2008): "Deep Habits and the Dynamic Effects of Monetary Policy Shocks," Economics Working Papers ECO2008/40, European University Institute.

Rigobon, R. (2003): "Identification Through Heteroskedasticity," *The Review of Economics and Statistics*, 85(4), 777–792.

Rubio-Ramirez, J. F., D. Waggoner, and T. Zha (2005): "Markov-switching structural vector autoregressions: theory and application," Working Paper 2005-27, Federal Reserve Bank of Atlanta.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), 1–48.

Sims, C. A. (2006b): "Making Macro Models Behave Reasonably," Working Papers 74, Princeton University, Department of Economics.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

Smets, F., and R. Wouters (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 97(3), 586–606.

Smith, A A, J. (1993): “Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions,” *Journal of Applied Econometrics*, 8(S), S63–84.

Uhlig, H. (1997): “Bayesian Vector Autoregressions with Stochastic Volatility,” *Econometrica*, 65(1), 59–74.

——— (2005a): “What are the effects of monetary policy on output? Results from an agnostic identification procedure,” *Journal of Monetary Economics*, 52(2), 381–419.

A Derivation of the posterior distribution of the BVAR

A.1 Prior distribution

$$
\begin{aligned}
vec(B)|\Sigma &\sim \mathcal{N}(vec(B_0), \Sigma \otimes N_0^{-1}) && \text{(A-1)}\\
\Sigma &\sim \mathcal{IW}(v_0 S_0, v_0) && \text{(A-2)}
\end{aligned}
$$

Σ is of size $m \times m$, N_0 of size $k \times k$, where $k = m * l$. The probability density function (p.d.f.) of $vec(B)$ is given by:

$$
\begin{aligned}
p(B|B_0, \Sigma, N_0) &= (2\pi)^{-mk/2} |\Sigma \otimes N_0^{-1}|^{-1/2} \\
&\quad exp\left[-\frac{1}{2}\left(vec(B) - vec(B_0)\right)' \left(\Sigma^{-1} \otimes N_0\right) \left(vec(B) - vec(B_0)\right)\right] \\
&= (2\pi)^{-mk/2} |\Sigma|^{-k/2} |N_0|^{m/2} exp\{-\frac{1}{2} tr\left[\Sigma^{-1}\left(B - B_0\right)' N_0 \left(B - B_0\right)\right]\}.
\end{aligned}
$$

The p.d.f. of Σ is defined as:

$$
p(\Sigma | v_0 S_0, v_0) = C_{IW}^{-1} |\Sigma|^{-\frac{1}{2}(v_0+m+1)} exp\left[-\frac{1}{2} tr\left(\Sigma^{-1} v_0 S_0\right)\right],
$$

where:

$$
C_{IW} = 2^{\frac{1}{2}v_0 m} \pi^{\frac{1}{4}m(m-1)} \prod_{i=0}^{m} \Gamma\left(\frac{v_0 + 1 - i}{2}\right) |S_0|^{-\frac{1}{2}v_0}.
$$

A.2 Likelihood

For

$$
vec(u) \sim \mathcal{N}(0, \Sigma \otimes I), \qquad \text{(A-3)}
$$

$$
p(Y|B,\Sigma) = (2\pi)^{-Tm/2} |\Sigma|^{-T/2} exp\{-\frac{1}{2} tr\left[\Sigma^{-1} (Y - XB)'(Y - XB)\right]\}. \qquad \text{(A-4)}
$$

The kernel can be rewritten as:

$$
\begin{aligned}
(Y - XB)'(Y - XB) &= (Y - XB - X\hat{B} + X\hat{B})'(Y - XB - X\hat{B} + X\hat{B}) \qquad \text{(A-5)}\\
&\quad (Y - X\hat{B})'(Y - X\hat{B}) + (B - \hat{B})' X' X (B - \hat{B}).
\end{aligned}
$$

A.3 Posterior

$$
\begin{aligned}
p(\Sigma, B|Y) &= C_{IW}^{-1}|\Sigma|^{-\frac{1}{2}(v_0+m+1)} exp\left[-\frac{1}{2}tr\left(\Sigma^{-1}v_0 S_0\right)\right] \quad \text{(A-6)}\\
&\times(2\pi)^{-mk/2}|\Sigma|^{-k/2}|N_0|^{m/2} exp\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(B-B_0\right)' N_0 \left(B-B_0\right)\right]\}\\
&\times(2\pi)^{-Tm/2}|\Sigma|^{-T/2} exp\{-\frac{1}{2}tr\left[\Sigma^{-1}(Y-X\hat{B})'(Y-X\hat{B})\right]\}\\
&\times exp\{-\frac{1}{2}tr\left[\Sigma^{-1}(B-\hat{B})'X'X(B-\hat{B})\right]\}
\end{aligned}
$$

Use the formula as stated in Leamer (1978)[16]:

$$
\begin{aligned}
(B-\hat{B})'X'X(B-\hat{B})\left(B-B_0\right)' N_0 \left(B-B_0\right) &= (B-B_T)'N_T(B-B_T) \quad \text{(A-7)}\\
&\times(B-B_0)'(X'X(N_T)^{-1}N_0)(B-B_0),
\end{aligned}
$$

where:

$$
\begin{aligned}
N_T &= N_0 + X'X\\
B_T &= N_T^{-1}(N_0 B_0 + X'X\hat{B})
\end{aligned}
$$

leads to:

$$
\begin{aligned}
p(\Sigma, B|Y) &= C_{IW}^{-1}|\Sigma|^{-\frac{1}{2}(v_0+m+1)} exp\left[-\frac{1}{2}tr\left(v_0\Sigma^{-1}S_0\right)\right] \quad \text{(A-8)}\\
&\times(2\pi)^{-mk/2}|\Sigma|^{-k/2}|N_0|^{m/2} exp\{-\frac{1}{2}tr\left[\Sigma^{-1}(B-B_0)'(X'X(N_T)^{-1}N_0)(B-B_0)\right]\}\\
&\times(2\pi)^{-Tm/2}|\Sigma|^{-T/2} exp\{-\frac{1}{2}tr\left[\Sigma^{-1}(Y-X\hat{B})'(Y-X\hat{B})\right]\}\\
&\times exp\{-\frac{1}{2}tr\left[\Sigma^{-1}(B-B_T)'N_T(B-B_T)\right]\}
\end{aligned}
$$

$$
\begin{aligned}
p(\Sigma, B|Y) &= C_{IW}^{-1}|\Sigma|^{-\frac{1}{2}(T+v_0+m+1)} \quad \text{(A-9)}\\
&\times exp\left[-\frac{1}{2}tr\left(\Sigma^{-1}\left(\frac{v_0}{v_T}S_0 + \frac{T}{v_T}\tilde{\Sigma} + \frac{1}{v_T}(B-B_0)'(X'X(N_T)^{-1}N_0)(B-B_0)\right)\right)\right]\\
&\times(2\pi)^{-m(T+k)/2}|\Sigma|^{-(T+k)/2} exp\{-\frac{1}{2}tr\left[\Sigma^{-1}(B-B_T)'N_T(B-B_T)\right]\}.
\end{aligned}
$$

[16] Appendix 1, T10

B Description and solution of the FTPL model

B.1 FTPL Model Setup

$$U_{F,t} = log(c_{F,t}) + log(m_{F,t}) \tag{A-10}$$

$$c_{F,t} + m_{F,t} + b_{F,t} + \tau_{F,t} = y_F + \frac{1}{\pi_{F,t}} m_{F,t-1} + \frac{R_{F,t-1}}{\pi_{F,t}} b_{F,t} \tag{A-11}$$

First-order conditions:

$$\frac{1}{R_{F,t}} = \beta_F \frac{1}{\pi_{F,t+1}} \tag{A-12}$$

$$m_{F,t} = c_F \frac{R_{F,t}}{R_{F,t} - 1} \tag{A-13}$$

Government budget constraint:

$$b_{F,t} + m_{F,t} + \tau_{F,t} = g_F + \frac{M_{F,t-1}}{P_{F,t}} + R_{F,t-1} \frac{B_{F,t-1}}{P_{F,t}} \tag{A-14}$$

Monetary authority:

$$R_{F,t} = \alpha_{F0} + \alpha_F \pi_{F,t} + \theta_{F,t} \tag{A-15}$$

$$\theta_{F,t} = \rho_{F1} \theta_{F,t-1} + \epsilon_{F1,t} \tag{A-16}$$

$$\epsilon_{F1,t} \sim N(0, \sigma_{F1}) \tag{A-17}$$

Fiscal authority:

$$\tau_{F,t} = \gamma_{F0} + \gamma b_{F,t-1} + \psi_{F,t} \tag{A-18}$$

$$\psi_{F,t} = \rho_{F2} \psi_{F,t-1} + \epsilon_{F2,t} \tag{A-19}$$

$$\epsilon_{F2,t} \sim N(0, \sigma_F 2) \tag{A-20}$$

B.2 Linearization

$$\bar{x}\hat{x}_t = \tilde{x}_t.$$

First equation:

$$\begin{aligned} R_{F,t} &= \alpha_{F0} + \alpha_F \pi_{F,t} + \theta_{F,t} \\ \pi_{F,t+1} &= \beta_F \alpha_{F0} + \beta_F \alpha_F \pi_{F,t} + \beta_F \theta_{F,t} \\ \tilde{\pi}_{F,t+1} &= \beta_F \alpha_f \tilde{\pi}_{F,t} + \beta_F \theta_{F,t} \end{aligned}$$

Second equation:

$$\begin{aligned}\bar{R}_F \hat{R}_{F,t} &= \frac{\bar{\pi}_F}{\beta_F} \hat{\pi}_{F,t+1} \\ \tilde{R}_{F,t} &= \frac{\tilde{\pi}_{F,t+1}}{\beta_F}\end{aligned}$$

$$\begin{aligned}m_{F,t} &= c_F \frac{R_{F,t}}{R_{F,t}-1} \\ \tilde{m}_{F,t} &= -\frac{c_F}{(\bar{R}_F-1)^2 \beta_F} \tilde{\pi}_{F,t+1}\end{aligned}$$

$$\begin{aligned}\tilde{m}_{F,t} &= -\frac{c_F}{(\bar{R}_F-1)^2 \beta_F} (\beta_F \alpha_F \tilde{\pi}_{F,t} + \beta_F \theta_{F,t}) \\ \tilde{m}_{F,t-1} &= -\frac{c_F \alpha_F}{(\bar{R}_F-1)^2} \tilde{\pi}_{F,t-1} - \frac{c_F}{(\bar{R}_F-1)^2} \theta_{F,t-1}\end{aligned}$$

$$\begin{aligned}b_{F,t} + m_{F,t} + \tau_{F,t} &= g_F + \frac{M_{F,t-1}}{P_{F,t}} + R_{F,t-1} \frac{B_{F,t-1}}{P_F, t} \\ \tilde{b}_{F,t} + \tilde{m}_{F,t} + \tilde{\tau}_{F,t} &= \frac{\tilde{m}_{F,t-1}}{\bar{\pi}_F} - \frac{\bar{m}_F}{\bar{\pi}_F^2} \tilde{\pi}_{F,t} + \frac{\bar{b}_F}{\bar{\pi}_F} \tilde{R}_{F,t-1} - \frac{\bar{R}_F \bar{b}_F}{\bar{\pi}_F^2} \tilde{\pi}_{F,t} + \frac{\bar{R}_F}{\bar{\pi}_F} \tilde{b}_{F,t-1}\end{aligned}$$

$$\begin{aligned}\tilde{b}_{F,t} - \frac{c_F \alpha_F}{(\bar{R}_F-1)^2} \tilde{\pi}_{F,t} + \frac{c_F \bar{R}_F}{(\bar{R}_F-1)\bar{\pi}_F^2} \tilde{\pi}_{F,t} + \frac{\bar{R}_F \bar{b}_F}{\bar{\pi}_F^2} \tilde{\pi}_{F,t} - \frac{c_F}{(\bar{R}_F-1)^2} \theta_{F,t} + \tilde{\tau}_{F,t} \\ = \frac{\tilde{m}_{F,t-1}}{\bar{\pi}_F} + \frac{\bar{b}_F}{\bar{\pi}_F} \tilde{R}_{F,t-1} + \frac{\bar{R}_F}{\bar{\pi}_F} \tilde{b}_{F,t-1}\end{aligned}$$

B.3 Simplifying the FTPL model

Define:

$$\begin{aligned}-\frac{c_F \alpha_F}{(\bar{R}_F-1)^2} + \frac{c_F \bar{R}_F}{(\bar{R}_F-1)\bar{\pi}_F^2} + \frac{\bar{R}_F \bar{b}_F}{\bar{\pi}_F^2} &= \frac{c_F}{(\bar{R}_F-1)} \left(-\frac{\alpha_F}{(\bar{R}_F-1)} + \frac{c_F}{\beta_F \bar{\pi}_F} \right) + \frac{\bar{b}_F}{\beta_F \bar{\pi}_F} = \varphi_{F1} \\ -\frac{c_F}{(\bar{R}_F-1)^2} &= \varphi_{F3}\end{aligned}$$

$$
\begin{aligned}
-\frac{1}{\bar{\pi}_F}\frac{c_F\alpha_F}{(\bar{R}_F-1)^2}+\frac{\bar{b}_F}{\bar{\pi}_F}\alpha_F &= -\frac{\alpha_F}{\bar{\pi}_F}\left[\frac{c_F}{(\bar{R}_F-1)^2}-\bar{b}_F\right]=-\varphi_{F2} \\
-\frac{1}{\bar{\pi}_F}\frac{c_F}{(\bar{R}_F-1)^2}+\frac{\bar{b}_F}{\bar{\pi}_F} &= -\frac{1}{\bar{\pi}_F}\left[\frac{c_F}{(\bar{R}_F-1)^2}-\bar{b}_F\right]=\frac{-\varphi_{F2}}{\alpha_F}=-\varphi_{F4}
\end{aligned}
$$

This yields:

$$
\tilde{b}_{F,t}+\varphi_{F1}\tilde{\pi}_{F,t}+\varphi_{F3}\theta_{F,t}-(\beta_F^{-1}-\gamma_F)\tilde{b}_{F,t-1}+\psi_{F,t}+\varphi_{F4}\theta_{F,t-1}+\varphi_{F2}\tilde{\pi}_{F,t-1}=0 \quad \text{(A-21)}
$$

$$
\tilde{b}_{F,t}+\varphi_{F1}\tilde{\pi}_{F,t}+\varphi_{F3}\theta_{F,t}+\psi_{F,t}=(\beta_F^{-1}-\gamma_F)\tilde{b}_{F,t-1}-\varphi_{F4}\theta_{F,t-1}-\varphi_{F2}\tilde{\pi}_{F,t-1} \quad \text{(A-22)}
$$

B.4 Calibration

Following Leeper (1991) the FTPL model is calibrated by setting:

$$
\begin{aligned}
\beta_F &= 0.99 \\
\bar{c}_F &= 0.75 \\
\frac{\bar{b}_F}{\bar{y}_F} &= 0.4 \\
\bar{\pi}_F &= 3.43 \\
\rho_{F1} &= 0.8 \\
\rho_{F2} &= 0 \\
\sigma_{F1} &= 0.2 \\
\sigma_{F2} &= 0.2
\end{aligned}
$$

C Solution of the Deep habits model

C.1 Steady state

$$
\begin{aligned}
R_H^{\star} &= 1/\beta_H && \text{(A-23)}\\
\bar{h}_H &= 0.3 && \text{(A-24)}\\
\bar{x}_H &= (1-\theta_H^d)\bar{h}_H && \text{(A-25)}\\
\bar{c}_H &= \bar{h}_H && \text{(A-26)}\\
\bar{\lambda}_H^c &= 1/((1-\theta_H^d)\eta_H) && \text{(A-27)}\\
\bar{\lambda}_H^y &= 1+(\theta_H^d\beta_H-1)\bar{\lambda}_H^c && \text{(A-28)}\\
\bar{w}_H &= \bar{\lambda}_H^y && \text{(A-29)}\\
\bar{\lambda}_H^h &= \bar{w}_H/(\bar{x}_H^{\sigma}\phi_H) && \text{(A-30)}\\
\gamma_H &= (\bar{x}_H^{-\sigma}\bar{w}_H - \bar{\lambda}_H^{\bar{h}_H})/\bar{h}_H^{\kappa} && \text{(A-31)}
\end{aligned}
$$

C.2 Loglinearized equations

$$\bar{x}_H\hat{x}_{H,t} = \bar{c}_H\hat{c}_{H,t} - \theta_H^d\bar{c}_H\hat{c}_{H,t-1} \qquad \text{(A-32)}$$

$$\gamma_H\bar{h}_H^{\kappa_H}\kappa_H\hat{h}_{H,t} = \bar{x}_H^{-\sigma_H}\bar{w}_H(-\sigma_H\hat{x}_{H,t}+\hat{w}_{H,t}) - \bar{\lambda}_H\hat{\lambda}_{H,t}^h \qquad \text{(A-33)}$$

$$\phi_H\bar{\lambda}_H^h\bar{h}_H\bar{x}_H^{\sigma}(\hat{\lambda}_{H,t}^h+\hat{h}_{H,t}+\sigma_H\hat{x}_{H,t})+\zeta_{Hw}(\hat{\pi}_{Hw,t}-\hat{\bar{\pi}}_{Hw,t}) = \bar{h}_H\bar{w}_H(\hat{h}_{H,t}+\hat{w}_{H,t})+\beta\zeta_{Hw}(\hat{\pi}_{Hw,t+1}-\hat{\bar{\pi}}_{Hw,t+1}) \qquad \text{(A-34)}$$

$$-\sigma_H\hat{x}_{H,t} = R_H^{\star}\hat{R}_{H,t} - \sigma\hat{x}_{H,t+1} - \hat{\pi}_{H,t+1} \qquad \text{(A-35)}$$

$$
\begin{aligned}
\hat{c}_{H,t} &= \hat{h}_{H,t} && \text{(A-36)}\\
\hat{\lambda}_{H,t}^y &= \hat{w}_{H,t} && \text{(A-37)}\\
\hat{h}_{H,t} &= \hat{y}_{H,t} && \text{(A-38)}
\end{aligned}
$$

$$\bar{\lambda}_H^y\hat{\lambda}_{H,t}^y + \bar{\lambda}_H^c\hat{\lambda}_{H,t}^c = \theta_H^d\beta_H\bar{\lambda}_H^c(-\sigma_H\hat{x}_{H,t+1}+\sigma_H\hat{x}_{H,t}+\hat{\lambda}_{H,t+1}^c) \qquad \text{(A-39)}$$

$$\eta_H\bar{\lambda}_H^c\bar{x}_H(\hat{\lambda}_{H,t}^c+\hat{x}_{H,t}) + \zeta_{Hp}(\hat{\pi}_{H,t}-\hat{\bar{\pi}}_{H,t}) = \bar{c}_H\hat{c}_{H,t} + \beta_H\zeta_{Hp}(\hat{\pi}_{H,t+1}-\hat{\bar{\pi}}_{H,t+1}) \qquad \text{(A-40)}$$

$$\hat{R}_{H,t} = \rho_{Hr}\hat{R}_{H,t-1} + (1-\rho_{Hr})(\alpha_{H\pi}\hat{\pi}_{H,t} + \alpha_{Hy}\hat{y}_{H,t}) + \epsilon_{H,t} \quad \text{(A-41)}$$

$$\hat{\bar{\pi}}_{H,t} = (1-\nu_{Hp})\hat{\pi}_{H,t-1} \quad \text{(A-42)}$$

$$\hat{\bar{\pi}}_{Hw,t} = (1-\nu_{Hw})\hat{\pi}_{Hw,t-1} \quad \text{(A-43)}$$

$$\bar{w}_H\hat{w}_{H,t} = \bar{w}_H\hat{w}_{H,t-1} + \hat{\pi}_{Hw,t} - \hat{\pi}_{H,t} \quad \text{(A-44)}$$

C.3 Calibration and estimation of the deep habits model

Following Ravn, Schmitt-Grohé, Uribe, and Uuskula (2008), calibrated values for the structural parameters are set as:

$$\begin{aligned}
R^\star_H &= 1.01 \\
\beta_H &= 1/R^\star_H \\
\phi_H &= 4 \\
\kappa_H &= 0.5 \\
\pi^\star_H &= 1 \\
\sigma_H &= 3 \\
\pi^\star_{Hw} &= 1 \\
\nu_{Hw} &= 0.96
\end{aligned}$$

Table 2: Prior distribution, Monte Carlo results and Posterior estimates of the structural parameters for the Deep habits model. Columns 1-4 specify the name and type of prior distribution with corresponding mean and standard deviation. Column 5 displays the parameter value for the data that was simulated from the DSGE model. Columns $mean_1$, std_1 give the Monte Carlo estimation results for the small parametrization of the DSGE model, columns $mean_2$, std_2 for the full parametrization. The last two columns display the estimation results from confronting the DSGE model with the data.

		Prior distribution		*Monte Carlo Experiment*					*Posterior distribution*	
Parameter	distribution	mean	std	sim	$mean_1$	std_1	$mean_2$	std_2	mean	std
θ_H^d	beta	0.5	0.2	0.85	0.84	0.01	0.84	0.01	0.72	0.12
η_H	normal	2	0.3	2.5	2.44	0.12	2.11	0.11	2.47	0.60
ζ_{Hw}	normal	40	5	40.9	40.21	1.23	40.60	0.55	42.50	15.87
ζ_{Hp}	normal	7	3	14.5	13.71	1.51	13.50	1.75	14.89	5.08
ν_{Hp}	normal	0.1	0.01	0	-	-	0.10	0.01	0.10	0.01
ρ_{Hr}	beta	0.5	0.2	0.7	-	-	0.67	0.03	0.81	0.07
$\alpha_{H\pi}$	beta	1.5	0.25	1.46	-	-	1.40	0.12	1.56	0.68
α_{Hy}	beta	0.125	0.1	0.1	-	-	0.05	0.02	0.01	0.25
σ_{Hr}	normal	2	8 *d.o.f.*	1	0.6	0.15	0.64	0.13	0.36	0.10

D Data description

The frequency of all data used is quarterly. The data ranges from 1955.1 to 2009.1. All series except the Fed Funds rate are in logs. GDP, personal consumption and real wages are transformed into per capita.

Nominal GDP: This series is *BEA NIPA Table 1.1.5. Gross Domestic Product.*

Private Consumption: This series is *BEA NIPA Table 1.1.5. Personal consumption expenditures.*

Wage: The wage rate is the series *COMPNFB, Nonfarm Business Sector: Compensation Per Hour* at the Federal Reserve Board of St. Louis' website

http://research.stlouisfed.org/fred2/series/COMPRNFB.

Interest Rate: This is the Federal Funds rate taken from

http://research.stlouisfed.org/fred2/series/FEDFUNDS.

Adjusted reserves: This is the adjusted monetary base given by the series *adjressl http://research.stlouisfed.org/fred2/series/ADJRESSL.*

PPIC: This series is *http://research.stlouisfed.org/fred2/series/PPICRM.*

Real GDP: This series is *BEA NIPA Table 1.1.6. Real Gross Domestic Product.*

Implicit GDP Deflator: The implicit GDP deflator is calculated as the ratio of **Nominal GDP** to **Real GDP**

Civilian Population: This is a quarterly measure for the population given by the respective average of the monthly values of the series *CNP16OV, Civilian Noninstitutional Population* at the Federal Reserve Board of St. Louis' website *http://research.stlouisfed.org/fred2/.* The numbers have been converted from thousands to billions.

E Figures

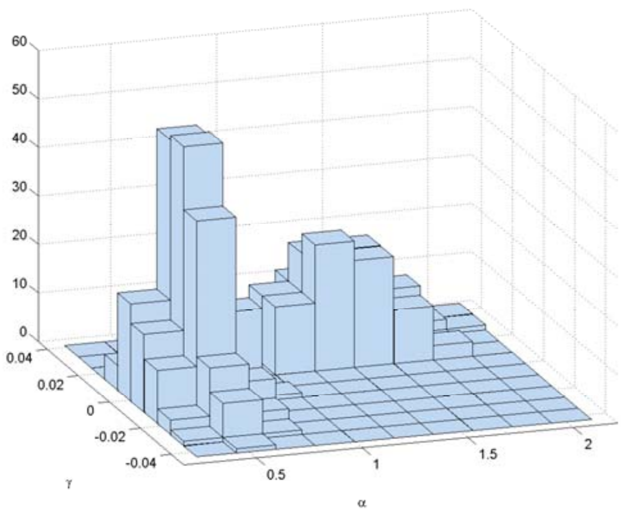

Figure 1: Prior distribution FTPL model

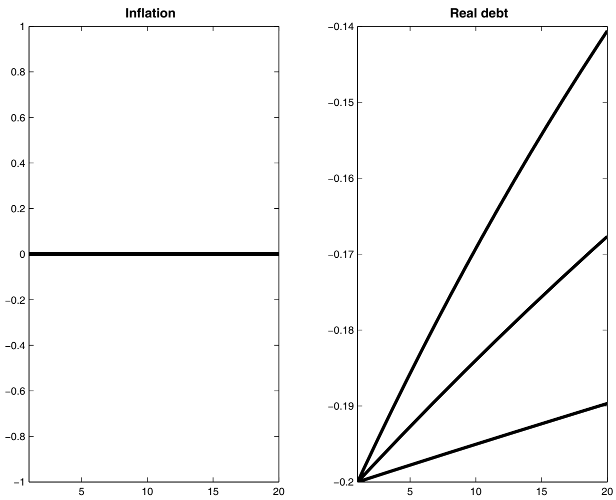

Figure 2: Prior Bayesian IRF for a fiscal policy shock regime I in the FTPL model

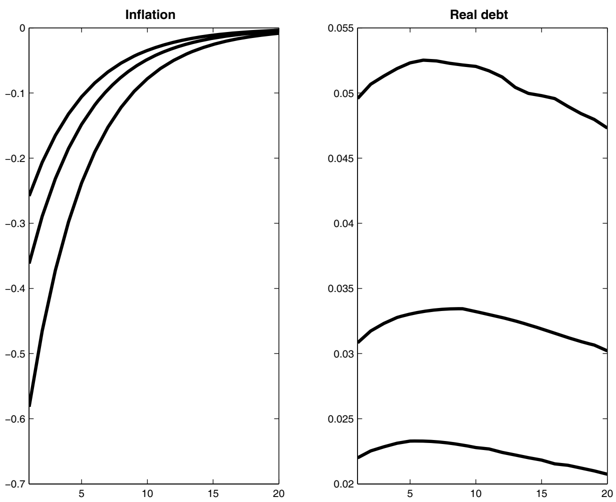

Figure 3: Prior Bayesian IRF for a monetary policy shock regime I in the FTPL model

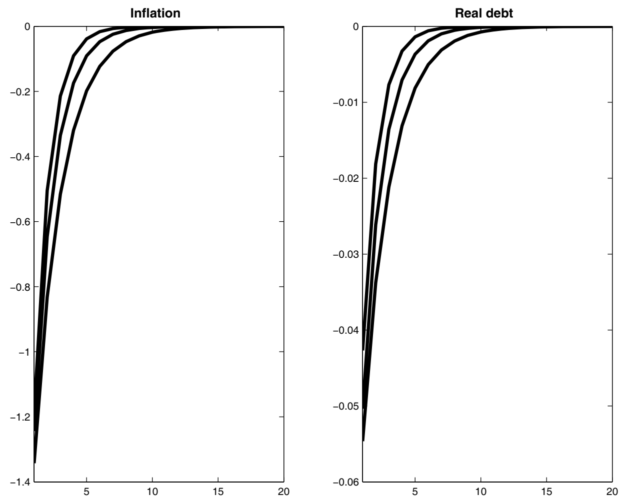

Figure 4: Prior Bayesian IRF for a fiscal policy shock regime II in the FTPL model

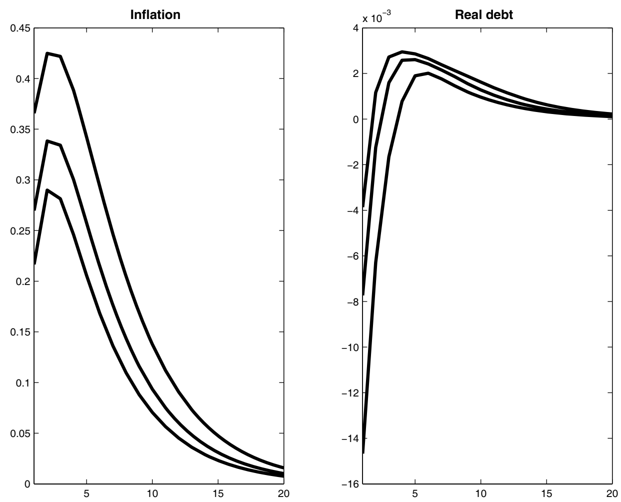

Figure 5: Prior Bayesian IRF for a monetary policy shock regime II in the FTPL model

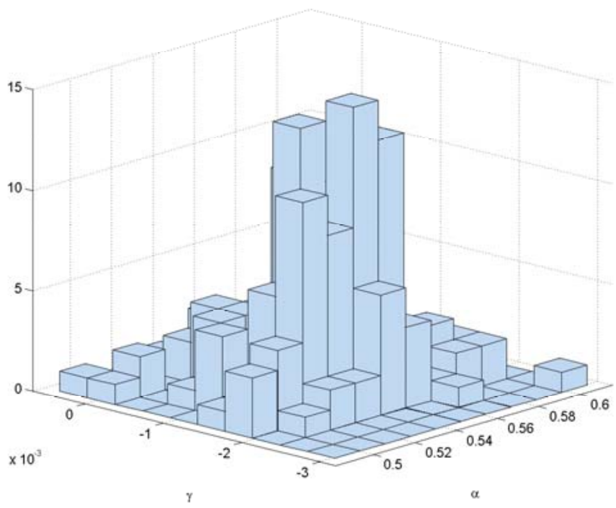

Figure 6: Posterior distribution FTPL model

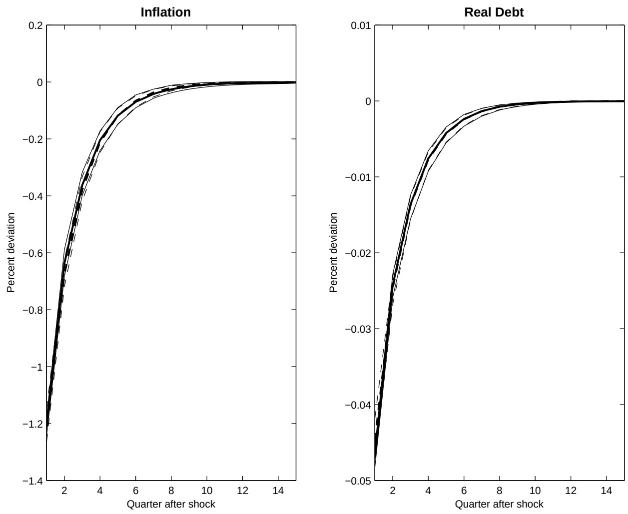

Figure 7: Estimated Bayesian IRF for a fiscal policy shock in the FTPL model: VAR model (black line) vs. DSGE model (dashed line).

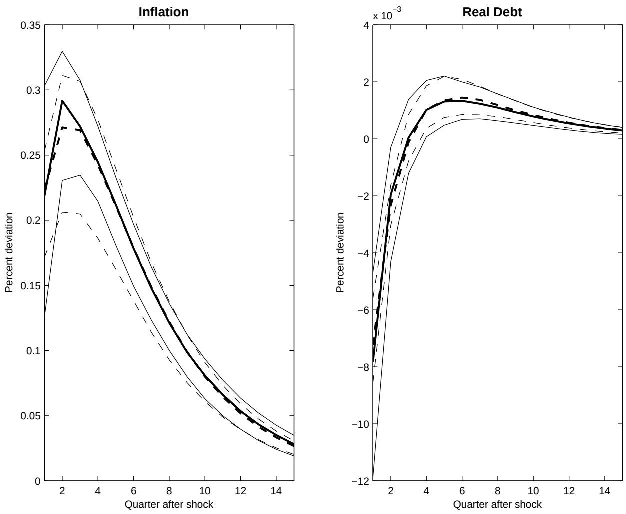

Figure 8: Estimated Bayesian IRF for a monetary policy shock in the FTPL model: VAR model (black line) vs. DSGE model (dashed line).

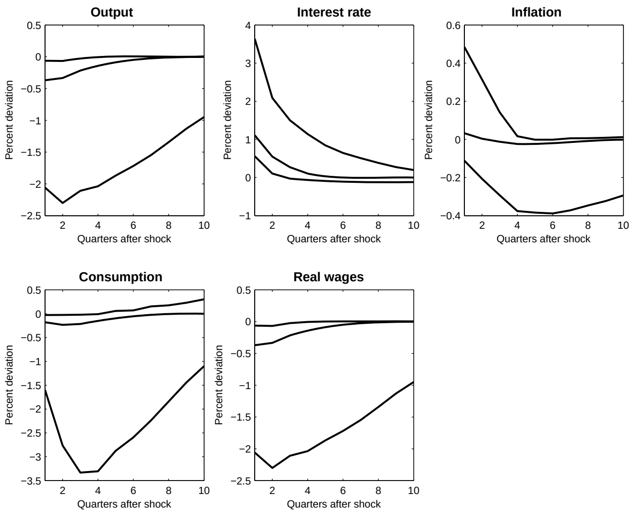

Figure 9: Impulse response function of the deep habits model drawing from the prior distribution of deep parameters (100 % probability bands).

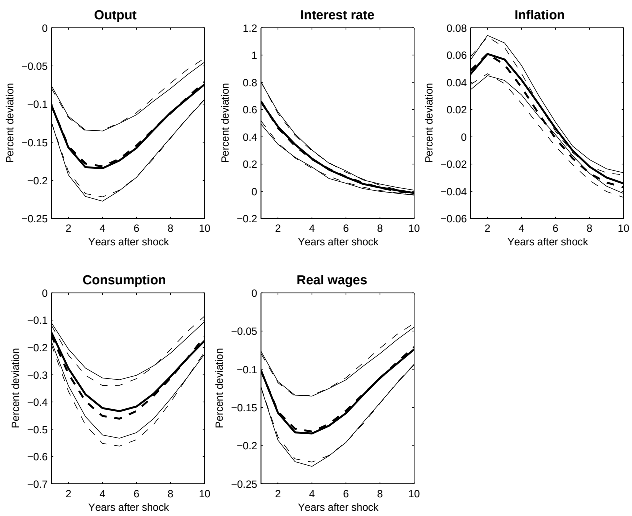

Figure 10: Impulse response functions of the deep habits model (dashed line) versus VAR model with simulated data (68 % probability bands).

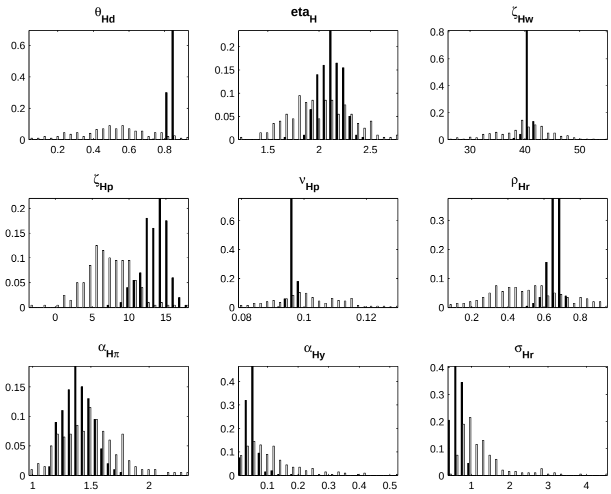

Figure 11: Prior distribution (white) vs. Posterior distribution (black). Monte-Carlo experiment deep habits model.

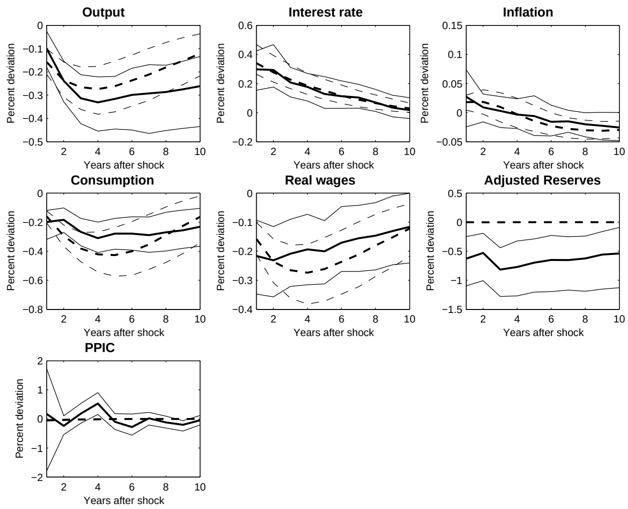

Figure 12: Posterior distribution of impulse response functions of the deep habits model (dashed line) versus VAR model (solid line) (68 % probability bands).

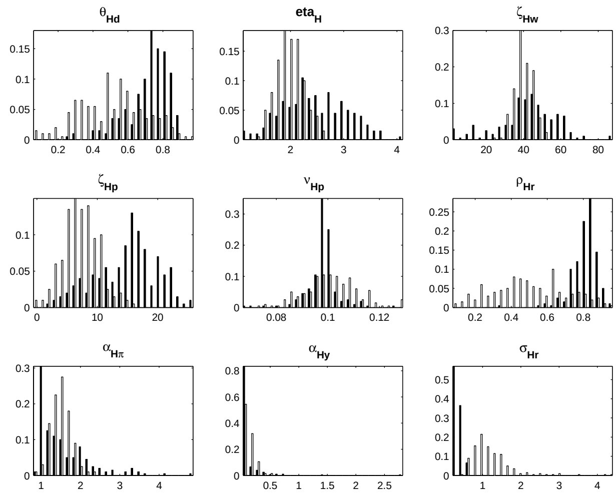

Figure 13: Prior distribution (white) vs. Posterior distribution (black). deep habits model.