Towards a monetary policy evaluation framework. *

Stéphane Adjemian† Matthieu Darracq Pariès‡ Stéphane Moyen§

May 13, 2008

Abstract

Advances in the development of Dynamic Stochastic General Equilibrium (DSGE) models towards medium-scale structural frameworks with satisfying data coherence have raised considerable interest among academia and policy circles over the recent years. The present paper intends to make a step forward in the development of structural policy tools: we explore within a unified framework the main approaches followed by the existing literature which *provocatively* assessed the *optimality* of historical monetary policy conduct. First, on US data over the Volcker-Greenspan sample, we perform a DGSE-VAR estimation of a medium-scale DSGE model very close to Smets and Wouters [2007] specification, where monetary policy is set according to a Ramsey-planner decision problem. Those results are then contrasted with the DSGE-VAR estimation of the same model featuring a Taylor-type interest rate rule. In doing so, we develop a policy evaluation framework which notably allows to assess whether optimal policy setting has been a good representation of historical monetary policy. Our results show in particular that the restrictions imposed by the welfare-maximizing Ramsey policy deteriorates the empirical performance with respect to a Taylor rule specification. However, it turns out that, along selected conditional dimensions, and in particular for productivity shocks, the optimal policy and the estimated Taylor rule deliver similar economic propagation.

Keywords: DSGE models, Optimal monetary policy, Bayesian estimation.
JEL classification: E4, E5, F4.

*First draft: February 2008. The views expressed are solely our own and do not necessarily reflect those of the European Central Bank.

†Université du Maine, GAINS & CEPREMAP.
‡European Central Bank.
§European Central Bank.

1 Introduction

Advances in the development of Dynamic Stochastic General Equilibrium (DSGE) models towards medium-scale structural frameworks with satisfying data coherence have raised considerable interest among academia and policy circles over the recent years. The ultimate goal of a challenging and abundant strand of literature has been to design analytical tools well-suited for monetary policy evaluation. The present paper intends to make a step forward in this direction by exploring in a unified framework the main approaches followed by the existing literature which *provocatively* assessed the *optimality* of historical monetary policy conduct.

Over the last decade, the quantitative and normative toolbox available to policy analysts has been expanded by promising research contributions. First, bayesian estimation techniques make it possible to estimate relatively large DSGE models. Smets and Wouters [2007] in particular have successfully brought to 7 macroeconomic series a closed-economy DSGE model for the US economy which could advantageously compare with vector autoregressions in terms of marginal data density and out-of-sample forecasts. In this model, monetary policy is specified as an interest rate feedback rule. At the same time, computational methods allow to easily derive optimal monetary policy concepts. The Ramsey approach to optimal monetary policy is calculated by formulating an infinite-horizon Lagrangian problem of maximizing the conditional aggregate welfare, subject to the full set of non-linear constraints forming the competitive equilibrium of the model. We solve the equilibrium conditions of the optimal allocation using second-order approximations to the policy function. Examples of Ramsey policy analysis in estimated closed-economy models can be found in Levin et al. [2005] for the US or Adjemian et al. [2007] for the euro area. Finally, a recent literature, led by the seminal work of Del Negro and Schorfheide [2004], has proposed an interesting metric to evaluate the potential misspecifications of DSGE models: the approach uses the DSGE model to shape the prior odds for a Bayesian VAR and provide an identification scheme consistent with the theoretical model. In this set-up, the optimal weight on the DSGE model for the BVAR priors as well as the comparison of impulse responses between the structural BVAR (or DSGE-VAR) and the DSGE constitute key dimensions to assess the validity of economic restrictions implied by the structural model.

In this paper, we conduct a DGSE-VAR estimation on US data of a medium-scale DSGE model very close to Smets and Wouters [2007] specification where monetary policy is set according to a Ramsey planner decision problem. Those results are then contrasted with the DSGE-VAR estimation of the same model with a Taylor rule specification (including terms on lagged inflation, lagged output gap and its first difference), for the later can be considered *ex ante* as the best-performing structural description of the data generating process. In doing so, we develop a policy evaluation framework which allows to investigate the various directions in which the literature has formed a normative assessment on historical monetary policy conduct.

Primarily, the paper provides a contribution on the estimation of structural models subject to the restriction that policy behaves optimally, in the vein of Salemi [2006], Dennis [2006] or Favero and Rovelli [2003]. In contrast to these studies which generally assume that the monetary authority minimizes a specified loss function, our approach explicitly tackles the welfare-maximizing monetary policy. Then, by allowing for a ranking of policies, including the fully optimal one, based on empirical criteria, we provide a consistent framework to pursue counterfactual analysis, soundly rooted in a *best-performing* description of the economy. Such a counterfactual approach to revealing the social optimality of monetary policy was initiated by the seminal contribution of Rotemberg and Woodford [1997]. Finally, the DSGE-VAR methodology used in this paper enables us to assess the *optimality* of historical monetary policy setting conditionally on certain type of economic disturbances. This relates to the literature which uses partial information inference from minimum distance techniques, in order to test the similarity of the macroeconomic transmission of technological shocks in particular, between DSGE models embedding optimal policy setting and structural VARs. Our approach notably improves upon the existing studies by investigating such a *conditional optimality* for a wider set of structural shocks.

Beyond the methodological contribution of the paper, our results concerning US monetary policy over the Volker-Greenspan period can be summarized as follows. The DSGE-VAR estimations suggests that the Taylor rule specification provides a better description of US data than the Ramsey model over the last two decades. A *provocative* interpretation of the relative fit of both models would conclude that Fed's policy has not been *optimally* conducted. At the same time, the deterioration in empirical performance coming from the restrictions imposed by the welfare-maximizing Ramsey policy is commensurate to the one obtained by removing the first difference output gap term in the interest rate rule. Furthermore, while the statistical inference supports the Taylor rule model, counterfactual analysis points to relatively modest welfare costs of such a policy compared with the optimal allocation. Such results should nonetheless be taken cautiously given the lack of robustness of welfare calculations. Finally, the comparison of impulse response functions in the DSGE-VAR and the DSGE for the Ramsey and the Taylor rule models brings a conditional perspective on their relative empirical relevance. The transmission of a productivity shock to the US economy is very similar both between the Ramsey and the Taylor rule models, and between each DSGE and its associated DSGE-VAR. This strong result echoes findings from the partial information literature supporting the view that the Fed's response to technological shock has been optimal. However, the conclusion does not hold for other type of disturbances like consumer preference shocks for example where the optimal policy delivers propagation mechanisms at odds with the transmission portrayed by its associated DSGE-VAR and the Taylor rule model.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 describes the estimation and reports the results. Section 4, 5 and 6 discuss respectively three dimensions of policy evaluation: the empirical fit of the Ramsey and Taylor rule DSGE-VARs, the welfare cost properties of the estimated Taylor rule and the comparison of impulse response functions between both policies and their associated DSGE-VARs. Section 7 revisits the previous results on the so called pre-Volcker data sample. Finally, section 8 concludes.

2 Summary of the theoretical model

The theoretical model underlying our policy analysis is extensively based on Smets and Wouters [2007]. The authors have provided a successful exercise regarding the ability of structural models to provide satisfactory empirical properties. Indeed, the sophistication of their modeling framework is guided by the need to match a high level of data coherence for the US economy.

The necessary frictions are well-known and have become a standard features of medium-scale DSGE models (see Christiano et al. [2005]): adjustment costs on investment and capacity utilization, habit persistence and staggered nominal wage and price contracts with partial indexation. Compared with their earlier work (see Smets and Wouters [2003] and Smets and Wouters [2005]), the authors specified a Kimball aggregator (see Kimball [1995]) in both labor and goods markets which improved the statistical inference of nominal rigidity in price and wage settings. Model steady state features a balanced growth path which imposes a common trend on output, consumption, investment and real wages. Finally, we retained a similar set of structural disturbances. The paper does not intend to make progress in the structural specification of Smets and Wouters [2007] but on the contrary, we restrain our policy evaluation to this exact structural framework which turns to be a useful benchmark for the empirical literature based on DSGE models.

In order to present a self containing paper, the main decision problems are reported below as well as the necessary notations related to the empirical exercise[1].

2.1 Households behavior

The economy is populated by a continuum of heterogenous infinitely-lived households. Each household is characterized by the quality of its labour services, $h \in [0,1]$. At time t, the in-

[1] Details regarding the full set of equilibrium conditions can be obtained from the authors upon request.

tertemporal utility function of a generic household h is

$$\mathcal{W}_t(h) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \varepsilon_{t+j}^b \left[\frac{\left(C_{t+j}(h) - \eta C_{t+j-1}(h)\right)^{1-\sigma_c}}{1-\sigma_c} \exp\left(\widetilde{L} \frac{(\sigma_c - 1)}{(1+\sigma_l)} L_{t+j}^{h\;1+\sigma_l} \right) \right] \tag{1}$$

Household h obtains utility from consumption of an aggregate index $C_t(h)$, relative to an internal habit depending on its past consumption, while receiving disutility from the supply of their homogenous labor L_t^h. Utility also incorporates a consumption preference shock ε_t^b. $\tilde{L}$ is a positive scale parameter.

Conversely to Smets and Wouters [2007], we assume *internal* habit formation. As we are more interested in the normative implications of nominal rigidities, we choose an habit formation mechanism that does not generate by itself a distortion affecting the welfare.

Each household h maximizes its intertemporal utility under the following budget constraint:

$$\begin{aligned} \frac{B_t(h)}{P_t R_t} + C_t(h) + I_t(h) = \frac{B_{t-1}(h)}{P_t} &+ \frac{(1-\tau_{w,t}) W_t^h L_t^h + A_t(h) + T_t(h)}{P_t} \\ &+ r_t^k u_t(h) K_{t-1}(h) - \Psi\left(u_t(h)\right) K_{t-1}(h) + \Pi_t(h) \end{aligned} \tag{2}$$

where P_t is an aggregate price index, $R_t = 1 + i_t$ is the one period ahead nominal interest factor, $B_t(h)$ is a nominal bond, $I_t(h)$ is the investment level W_t^h is the nominal wage, $T_t(h)$ and $\tau_{w,t}$ are government transfers and time-varying labor tax, and

$$r_t^k u_t(h) K_{t-1}(h) - \Psi\left(u_t(h)\right) K_{t-1}(h) \tag{3}$$

represents the return on the real capital stock minus the cost associated with variations in the degree of capital utilization. The income from renting out capital services depends on the level of capital augmented for its utilization rate. The cost (or benefit) Ψ is an increasing function of capacity utilization and is zero at steady state, $\Psi(u^\star) = 0$. $\Pi_t(h)$ are the dividend emanating from monopolistically competitive intermediate firms. Finally $A_t(h)$ is a stream of income coming from state contingent securities and equating marginal utility of consumption across households $h \in [0, 1]$.

In choosing the capital stock, investment and the capacity utilization rate households take into account the following capital accumulation equation:

$$K_t = (1-\delta) K_{t-1} + \varepsilon_t^I \left[1 - S\left(\frac{I_t}{I_{t-1}} \right) \right] I_t \tag{4}$$

where $\delta \in (0, 1)$ is the depreciation rate, S is a non negative adjustment cost function such that $S\,(1) = 0$ and ε_t^I is an efficiency shock on the technology of capital accumulation.

In equilibrium, households choices in terms of consumption, hours, bond holdings, investment and capacity utilization are identical.

Thereafter, the functional forms used for the adjustment costs on capacity utilization and investment are given by $\Psi(X) = \frac{r^{k\star}}{\varphi}\left(\exp\left[\varphi\left(X-1\right)\right]-1\right)$ and $S\left(x\right) = \phi/2\ \left(x-1\right)^2$.

2.2 Labor supply and wage setting

Intermediate goods producers make use of a labor input L_t^D produced by a segment of labor packers. Those labor packers operate in a competitive environment and aggregate a continuum of differentiated labor services $L_t(i)$, $i \in [0,1]$ using a Kimball [1995] technology. The Kimball aggregator is defined by

$$\int_0^1 H\left(\frac{L_t(i)}{L_t^D};\theta_w,\psi_w\right)\mathrm{d}i = 1 \tag{5}$$

where as in Dotsey and King [2005], we consider the following functional form:

$$\begin{aligned} H\left(\frac{L_t(i)}{L_t^D}\right) &= \frac{\theta_w}{(\theta_w(1+\psi_w)-1)}\left[\left(1+\psi_w\right)\frac{L_t(i)}{L_t^D}-\psi_w\right]^{\frac{\theta_w(1+\psi_w)-1}{\theta_w(1+\psi_w)}} \\ &\quad -\left[\frac{\theta_w}{(\theta_w(1+\psi_w)-1)}-1\right] \end{aligned} \tag{6}$$

This function, where the parameter ψ_w determines the curvature of the demand curve, has the advantage that it reduces to the standard Dixit and Stiglitz [1977] aggregator under the restriction $\psi_w = 0$.

The differentiated labor services are produced by a continuum of unions which transform the homogeneous household labor supply. Each union is a monopoly supplier of a differentiated labour service and sets its wage on a staggered basis, paying households the nominal wage rate W_t^h. Every period, any union faces a constant probability $1-\alpha_w$ of optimally adjusting its nominal wage, say $W_t^*(i)$, which will be the same for all suppliers of differentiated labor services. We denote thereafter w_t the aggregate real wage that intermediate producers pay for the labor input provided by the labor packers and w_t^* the real wage claimed by re-optimizing unions.

When they cannot re-optimize, wages are indexed on past inflation and steady state inflation according to the following indexation rule:

$$W_t(i) = \left[\pi_{t-1}\right]^{\xi_w}\left[\pi^\star\right]^{1-\xi_w}W_{t-1}(i) \tag{7}$$

with $\pi_t = \frac{P_t}{P_{t-1}}$ the gross rate of inflation. Taking into account that they might not be able to choose their nominal wage optimally in a near future, $W_t^*(i)$ is chosen to maximize their intertemporal profit under the labor demand from labor packers. Unions are subject to a time-varying tax rate $\tau_{w,t}$ which is affected by an i.i.d shock defined by $1-\tau_{w,t} = (1-\tau_w^\star)\,\varepsilon_t^w$. The recursive formulation of the aggregate wage setting is exposed in the appendix.

2.3 Producers behavior

Final producers are perfectly competitive firms producing an aggregate final good Y_t that may be used for consumption and investment. This production is obtained using a continuum of differentiated intermediate goods $Y_t(z),\ z \in [0,1]$ with the Kimball [1995] technology. Here again, the Kimball aggregator is defined by

$$\int_0^1 G\left(\frac{Y_t(z)}{Y_t};\theta_p,\psi\right)\mathrm{d}z = 1 \tag{8}$$

with

$$\begin{aligned} G\left(\frac{Y_t(z)}{Y_t}\right) &= \frac{\theta_p}{(\theta_p(1+\psi)-1)}\left[(1+\psi)\,\frac{Y_t(z)}{Y_t}-\psi\right]^{\frac{\theta_p(1+\psi)-1}{\theta_p(1+\psi)}} \\ &\quad -\left[\frac{\theta_p}{(\theta_p(1+\psi)-1)}-1\right]. \end{aligned} \tag{9}$$

The representative final good producer maximizes profits $P_tY_t - \int_0^1 P_t(z)Y_t(z)\mathrm{d}z$ subject to the production function, taking as given the final good price P_t and the prices of all intermediate goods.

In the intermediate goods sector, firms $z \in [0,1]$ are monopolistic competitors and produce differentiated products by using a common Cobb-Douglas technology:

$$Y_t(z) = \varepsilon_t^a\left(u_tK_{t-1}(z)\right)^\alpha\left[\gamma^tL^D(z)\right]^{1-\alpha} - \gamma^t\Omega \tag{10}$$

where ε_t^a is an exogenous productivity shock, $\Omega > 0$ is a fixed cost and γ is the trend technological growth rate. A firm z hires its capital, $\widetilde{K}_t(z) = u_tK_{t-1}(z)$, and labor, $L_t^D(z)$, on a competitive market by minimizing its production cost. Due to our assumptions on the labor market and the rental rate of capital, the real marginal cost is identical across producers. We introduce a time varying tax on firm's revenue is affected by an i.i.d shock defined by $1-\tau_{p,t} = \left(1-\tau_p^\star\right)\varepsilon_t^p$.

In each period, a firm z faces a constant (across time and firms) probability $1-\alpha_p$ of being able to re-optimize its nominal price, say $P_t^*(z)$. If a firm cannot re-optimize its price, the nominal price evolves according to the rule $P_t(z) = \pi_{t-1}^{\xi_p}\left[\pi^\star\right]^{(1-\xi_p)}P_{t-1}(z)$, *ie* the nominal price is indexed on past inflation and steady state inflation. In our model, all firms that can re-optimize their price at time t choose the same level, denoted p_t^* in real terms.

The first order condition associated with the maximization of the intertemporal profit can be expressed in a recursive form as shown in the appendix.

2.4 Government

Public expenditures $G^{\star}$ are subject to random shocks ε_t^g. The government finances public spending with labor tax, product tax and lump-sum transfers:

$$P_t G^{\star} \gamma^t \varepsilon_t^g - \tau_{w,t} W_t L_t - \tau_{p,t} P_t Y_t - P_t T_t = 0 \tag{11}$$

The government also controls the short term interest rate R_t. In the Taylor rule version of the model, the monetary authority follows an interest rate feedback rule which incorporates terms on lagged inflation, lagged output gap and its first difference. The output gap is defined as the log-difference between actual and flexible-price output. The reaction function also incorporates a non-systematic component ε_t^r. This specification is the same as in Smets and Wouters [2007].

Written in deviation from the steady state, the interest rule used in the estimation has the form:

$$\hat{R}_t = \rho \hat{R}_{t-1} + (1-\rho)\left[r_\pi \hat{\pi}_{t-1} + r_y \hat{y}_{t-1}\right] + r_{\Delta y} \Delta \hat{y}_t + \log\left(\varepsilon_t^r\right) \tag{12}$$

where a hat over a variable denotes log-deviation of that variable from its deterministic steady-state level.

2.5 Market clearing conditions

Market clearing condition on goods market is given by:

$$Y_t = C_t + I_t + G^{\star} \varepsilon_t^g + \Psi\left(u_t\right) K_{t-1} \tag{13}$$

$$\Delta_{pk,t} Y_t = \varepsilon_t^a \left(u_t K_{t-1}\right)^\alpha \left(\gamma^t L_t^D\right)^{1-\alpha} - \gamma^t \Omega \tag{14}$$

with $\Delta_{pk,t}$ is a price dispersion index whose dynamics is presented in the appendix.

Equilibrium in the labor market implies that

$$\Delta_{wk,t} L_t^D = L_t \tag{15}$$

with $L_t^D = \int_0^1 L_t^D(z) dz$ and $L_t = \int_0^1 L_t^h dh$. The dynamics of the wage dispersion index $\Delta_{wk,t}$ is also described in the appendix.

Finally, the aggregate conditional welfare is defined by

$$\mathcal{W}_t = \int_0^1 \mathcal{W}_t(h) dh \tag{16}$$

2.6 Ramsey equilibrium

We define the Ramsey policy as the monetary policy under commitment which maximizes the intertemporal household's aggregate welfare $\mathcal{W}_t$, subject to the competitive equilibrium conditions and the constraint $R_t \geq 1, \forall t > -\infty$, given the exogenous stochastic processes ε_t^a, ε_t^b, ε_t^I, ε_t^g, ε_t^w, ε_t^p, ε_t^r, values of the state variables dated $t < 0$, and values of the Lagrange multipliers associated with the constraints dated $t < 0$.

The Ramsey policy is therefore computed by formulating an infinite-horizon Lagrangian problem of maximizing the conditional expected social welfare subject to the full set of non-linear constraints forming the competitive equilibrium of the model. The first order conditions to this problem are obtained using symbolic Matlab procedures.

As it is common in the optimal monetary policy literature (see for example Khan et al. [2003] and Schmitt-Grohe and Uribe [2005]), we assume a particular recursive formulation of the policy commitment labeled by Woodford [2003] as optimality *from a timeless perspective*. This imposes that the policy rule which is optimal in the latter periods is also optimal in the initial period and avoids the problem of finding initial conditions for the lagrange multipliers, which are now endogenous and given by their steady state values.

Since we are mainly interested in comparing the macroeconomic stabilization performances of different monetary policy regimes, we assume a fiscal intervention, namely subsidies on labor and goods markets, to offset the first order distortions caused by the presence of monopolistic competition in the markets. This ensure that the steady state is efficient, and that the flexible price equilibrium is Pareto optimal. Note that those constraints can be easily relaxed with our methodology but are imposed in order to better understand the stabilization properties of the Ramsey policy.

To handle the Zero Lower Bound constraint under the Ramsey allocation, $R_t \geq 1$, and to avoid the associated computational burden, we simply follow Woodford [2003] by introducing in the households welfare a quadratic term penalizing the variance of the nominal interest rate:

$$\mathcal{W}_t^{IR} = \mathcal{W}_t - \lambda_r \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \left(R_{t+j} - R^\star \varepsilon_{t+j}^r \right)^2 \tag{17}$$

where $R^\star$ is the steady state nominal interest gross rate, λ_r is the weight attached to the cost on nominal interest rate fluctuations and ε_t^r represents, as in the Taylor rule specification, a monetary policy shock.

3 DSGE-VAR estimations

In section 3.2, we present the estimation of two DSGE-VARs, one based on a DSGE with optimal monetary policy and the other with a Taylor rule specified as in Smets and Wouters [2007]. We closely follow the econometric approach used by Del Negro et al. [2007] who estimated a medium-scale closed-economy model on US data. A description of the DSGE-VAR methodology is provided in the Appendix.

In a nutshell, Del Negro and Schorfheide [2004] build the priors of a BVAR model from a DSGE model and evaluate the optimal weight of the DSGE priors. Their approach relies on a finite order VAR representation of the DSGE but the error of approximation should be relatively minor, at least with a reasonably larger lag length in the VAR. The posterior density is obtained from the likelihood function by augmenting the sample with artificial data generated by the DSGE model. The size of the artificial sample, $\mathcal{T}$ relative to the data sample T , defines the weight of the prior information relative to the likelihood. Let us denote $\lambda_{DSGE} = \frac{\mathcal{T}}{T}$. A crucial issue is to choose the *optimal* weight, λ_{DSGE}, of the DSGE prior in the BVAR model. An *optimal* high value of λ_{DSGE} means that the DSGE model imposes useful restrictions to improve the (in sample) predictive properties of the BVAR model. Conversely, a low value of λ_{DSGE} indicates that a minimal use of the DSGE restrictions on the priors of the BVAR is preferred, therefore casting doubts on the coherence of the DSGE model with the data.

The exogenous shocks can be divided in three categories [2]:

1. Efficient shocks: AR(1) shocks on technology ϵ_t^a, investment ϵ_t^I, public expenditures ϵ_t^g and consumption preferences ϵ_t^b.

2. Inefficient shocks: ARMA(1,1) shocks on price markups ϵ_t^p, and wage markups ϵ_t^w.

3. Policy shocks: AR(1) shock on short term interest rates ϵ_t^r.

Given the strict identification scheme used in the DSGE-VAR, we limited the number of shocks to be equal to the number of observed variables. Under such a configuration, the Ramsey equilibrium would be subject to a stochastic singularity problem at the estimation stage as it does not feature a Taylor rule residual shock. Therefore we allowed for some type of *monetary policy shock* in the Ramsey allocation through the penalty term on interest rate volatility introduced in the welfare function. This shock is, as its Taylor rule counterpart, intended to capture the historical dynamics of the policy instrument missed by the specified reaction function. Nonetheless, they are not fully equivalent. In the Ramsey model, the interest rate shock is strictly isomorphic

[2] All the AR(1) processes are written as: $\log(\varepsilon_t^x) = \rho_x \log(\varepsilon_{t-1}^x) + \epsilon_t^x$ where $\epsilon_t^x \sim \mathcal{N}(0, \sigma_{\varepsilon^x})$. ARMA(1,1) are of the form $\log(\varepsilon_t^x) = \rho_x \log(\varepsilon_{t-1}^x) - \eta_x \epsilon_{t-1}^x + \epsilon_t^x$.

to a risk premium shock à la Smets and Wouters [2007], in the same model where the interest rate is replaced by its deviation from the interest rate shock[3].

As in Smets and Wouters [2007], we introduced a correlation between the government spending shock and the productivity shock, $\rho_{a,g}$. But differently from them, we allowed for a correlation between preference shocks and external risk premium shocks, $\rho_{b,I}$, essentially to match the correlation between consumption and investment present in the data. The authors used instead a risk premium shock affecting consumer financing and acting as a common disturbance for the Euler and the Tobin's Q equations. From an empirical perspective, both specifications deliver similar outcome. We preferred to keep the household preference shock specification in order to have an efficient demand shock which may imply differentiated stabilization properties under the Ramsey policy and the Taylor rule (see Adjemian et al. [2007]).

3.1 Data

We consider 7 key macro-economic quarterly time series: output, consumption, investment, hours worked, real hourly wages, GDP deflator inflation rate and 3 month short-term interest rate and we use the Volker-Greenspan sample starting from 1983q1 to 2007q3. As it is usually done in the literature, we excluded the beginning of the 80's which were characterized by non-borrowed reserves targeting. US series come from the BEA for GDP, consumption, investment and nominal compensation of employees. The GDP deflator is used to compute real consumption, real investment and real compensation. Individual hours are taken from the BLS for the non-farm business sector and are combined with the Civilian Employment data to compute aggregate hours. The real aggregate variables are then expressed per capita, dividing by the population over 16. The interest rate is the Federal Funds Rate. For the estimation, we use the quarterly growth rates of real variables, the quarterly inflation rate and the quarterly interest rate, in percent.

As in Smets and Wouters [2007], the transformed data are not demeaned since the model features non-zero steady state values for such variables: trend productivity growth γ captures the common mean of GDP, consumption, investment and real wage growth; $\overline{L}$ is a level shift that we allow between the observers level of hours and the model consistent one; $\overline{\pi}$ is the steady state inflation rate which controls for the GDP deflator inflation rate mean; and finally, we also estimate the preference rate $r_\beta = 100(1/\beta - 1)$ which, combined with $\overline{\pi}$, pins down the mean of the nominal interest rate.

Section 7 also reports estimations on the pre-Volker sample which goes from 1966Q1 to 1979Q3.

[3]For an illustration, see the derivation of the optimal rule in Giannoni and Woodford [2003].

All estimations are initialized using a presample period of 20 quarters and we choose four lags in the DSGE-VAR representation.

Table 1: **Prior Distributions**

Parameter	Distribution	Mean	Std. dev.
σ_c	Normal	1	0.375
h	Beta	0.7	0.1
σ_l	Normal	2	0.75
ϕ	Gamma	0.2	0.1
φ	Normal	4.000	1.5
α_p, α_w	Beta	0.5	0.1
ξ_p, ξ_w	Beta	0.5	0.15
α	Normal	0.3	0.05
μ_p	Normal	1.25	0.2
r_β	Normal	0.25	0.1
γ	Normal	0.4	0.1
$\overline{L}$	Normal	0	2
$\overline{\pi}$	Gamma	0.62	0.1
λ_r	Gamma	0.2	0.15
r_π	Normal	1.500	0.100
ρ	Beta	0.750	0.100
r_y	Gamma	0.125	0.050
$r_{\Delta y}$	Gamma	0.063	0.050
$\rho_{b,I}$	Uniform		
$\rho_{a,g}$	Uniform		
ρ_a, ρ_b, ρ_g, ρ_l, ρ_I	Beta	0.50	0.2
ρ_r,ρ_p, ρ_w, η_p, η_w	Beta	0.50	0.2
σ_{ε^a}, σ_{ε^b} ,σ_{ε^g}	Uniform		
σ_{ε^I},σ_{ε^p},σ_{ε^w} ,σ_{ε^r}	Uniform		

3.2 Prior and Posterior parameter distributions

Like in Smets and Wouters [2007], some parameters are treated as fixed in the estimation. The depreciation rate of the capital stock is set at 0.025 and the share of government spending in output at 18%. The steady state labor market markup is fixed at 1.5 and we chose curvature parameters of the Kimball aggregators of 10.

The prior distributions for the structural parameters are also similar to Smets and Wouters [2007] and are reported in Table 1. The main differences relate to the choice of uniform priors for the standard deviations of the exogenous shocks. Concerning the parameter controlling the welfare penalty of interest rate fluctuations in the Ramsey problem, λ_r, we used a prior gamma

Table 2: **Posterior parameter estimates of Ramsey and Taylor rule DSGE-VARs for the Volker-Greenspan sample.**

	Ramsey			**Taylor**			**Ramsey**			**Taylor**		
	DSGE-VAR estimation			*DSGE-VAR estimation*			*DSGE estimation*			*DSGE estimation*		
	Mode	$\mathcal{I}_1$	$\mathcal{I}_2$	Mode	$\mathcal{I}_1$	$\mathcal{I}_2$	Mode	$\mathcal{I}_1$	$\mathcal{I}_2$	Mode	$\mathcal{I}_1$	$\mathcal{I}_2$
σ_c	0.86	0.65	1.12	0.84	0.63	1.09	1.28	0.97	1.61	0.87	0.67	1.12
η	0.66	0.51	0.77	0.58	0.46	0.68	0.67	0.47	0.80	0.54	0.45	0.63
σ_l	1.17	0.55	2.10	1.64	0.84	2.90	2.22	1.22	3.50	2.22	1.35	3.35
ϕ	4.98	3.14	7.08	4.79	3.08	7.18	5.61	3.63	7.80	5.46	3.66	7.42
φ	0.60	0.37	0.81	0.66	0.40	0.82	0.77	0.59	0.89	0.71	0.50	0.84
α_p	0.67	0.58	0.74	0.68	0.58	0.77	0.63	0.54	0.72	0.71	0.62	0.80
ξ_p	0.35	0.15	0.60	0.28	0.13	0.50	0.36	0.18	0.62	0.44	0.22	0.66
α_w	0.52	0.34	0.69	0.74	0.57	0.85	0.38	0.25	0.50	0.57	0.45	0.71
ξ_w	0.48	0.25	0.74	0.50	0.25	0.74	0.61	0.35	0.82	0.61	0.35	0.82
α	0.13	0.09	0.17	0.12	0.09	0.16	0.15	0.12	0.18	0.13	0.10	0.15
μ_p	1.20	1.09	1.39	1.39	1.24	1.53	1.19	1.11	1.29	1.36	1.24	1.49
r_β	0.17	0.04	0.30	0.19	0.06	0.32	0.20	0.07	0.35	0.25	0.12	0.36
γ	0.44	0.35	0.53	0.44	0.36	0.52	0.44	0.41	0.46	0.42	0.39	0.45
$\overline{L}$	0.90	-0.12	1.85	0.90	-0.10	1.88	-1.31	-3.54	0.88	0.16	-2.00	1.99
$\overline{\pi}$	0.60	0.50	0.72	0.62	0.52	0.73	0.64	0.49	0.78	0.64	0.51	0.82
λ_r	0.18	0.06	0.52	-	-	-	0.79	0.25	4.92			
r_π		-	-	1.60	1.22	1.98	-	-	-	2.01	1.70	2.33
ρ		-	-	0.84	0.78	0.88	-	-	-	0.86	0.82	0.88
r_Y		-	-	0.16	0.09	0.24	-	-	-	0.06	0.03	0.10
$r_{\Delta Y}$		-	-	0.16	0.10	0.22	-	-	-	0.20	0.15	0.24
$\rho_{b,I}$	0.76	0.26	1.80	0.22	0.09	0.88	0.81	0.27	1.94	0.18	0.08	0.47
$\rho_{a,g}$	2.70	1.62	3.75	2.42	1.29	3.67	2.22	1.26	3.19	2.02	1.05	3.11
ρ_a	0.98	0.78	1.00	0.91	0.67	0.97	0.92	0.85	0.97	0.93	0.87	0.97
ρ_b	0.37	0.14	0.58	0.80	0.54	0.90	0.22	0.07	0.56	0.84	0.69	0.91
ρ_g	0.97	0.66	0.99	0.88	0.65	0.99	0.97	0.94	0.98	0.97	0.94	0.99
ρ_I	0.33	0.12	0.52	0.35	0.15	0.58	0.64	0.50	0.78	0.61	0.47	0.76
ρ_p	0.43	0.20	0.73	0.14	0.02	0.36	0.84	0.72	0.92	0.81	0.64	0.94
η_p	0.55	0.30	0.85				0.58	0.35	0.74	0.67	0.42	0.82
ρ_w	0.41	0.20	0.71	0.29	0.07	0.66	0.98	0.96	0.99	0.99	0.98	1.00
η_w	0.49	0.32	0.70	0.41	0.22	0.70	0.67	0.47	0.80	0.90	0.79	0.96
ρ_r	0.94	0.84	0.98	0.30	0.15	0.44	0.96	0.93	0.98	0.28	0.16	0.43
σ_{ε^a}	0.33	0.27	0.40	0.31	0.26	0.37	0.39	0.34	0.44	0.37	0.33	0.42
σ_{ε^b}	0.94	0.67	1.37	0.98	0.72	1.36	1.59	1.01	2.46	1.31	1.04	1.66
σ_{ε^g}	1.53	1.26	1.82	1.63	1.36	1.92	2.16	1.91	2.43	2.15	1.91	2.45
σ_{ε^I}	4.26	2.55	6.48	3.91	2.28	6.24	4.94	3.00	7.08	3.97	2.65	6.38
σ_{ε^p}	0.13	0.11	0.17	0.11	0.08	0.14	0.14	0.11	0.18	0.13	0.11	0.15
σ_{ε^w}	0.08	0.07	0.10	0.09	0.08	0.11	0.13	0.11	0.14	0.12	0.10	0.14
σ_{ε^r}	0.41	0.32	0.51	0.38	0.31	0.48	0.52	0.41	0.71	0.43	0.37	0.52
λ_{DSGE}	1.30	1.02	1.70	1.57	1.16	2.18	-	-	-	-	-	-
$P_\lambda(\mathcal{Y})$	-398.1			-393.8			-474.3			-442.0		

$\mathcal{G}(0.2, 0.15)$. We choose the prior mean so that, given the structural parameter estimates for the *DSGE-VAR-Taylor* model, the *DSGE-VAR-Ramsey* implies an unconditional variance of the nominal interest rate close to the *DSGE-VAR-Taylor* one. Finally, this prior applies to λ_r once rescaled by the coefficient on the inflation term that would appear in a quadratic approximation of the welfare, albeit in a simpler version of the model.

Del Negro and Schorfheide choose the value of λ_{DSGE} that maximizes the marginal density. They estimate a limited number of DSGE-VAR models with different values of λ_{DSGE}. For each model they also estimate the marginal density and select the model (*ie* the value of λ_{DSGE}) with highest marginal density. In the present paper, we estimate directly λ_{DSGE} as another parameter, instead of doing a loop over the values of this parameter[4]. We chose a uniform prior distribution for λ_{DSGE}.

The posterior parameter estimates in the Taylor rule and the Ramsey policy, for both the DSGE-VAR and the direct DSGE estimation approach, are presented in Table 2. Note that we replaced the ARMA(1,1) specification for the price markup shock in the Taylor rule DSGE-VAR by an AR(1) process which was performing better in the estimation.

Overall, in the DSGE-VAR estimation, the behavioral parameters estimates are not strongly different between both models. This result brings some reassurance that the structural inference made on aggregate supply and demand curves in our modeling framework are not excessively sensitive to monetary policy specification. Few exceptions are nonetheless worth noticing. Regarding preferences, the labor supply elasticity is lower in the Ramsey DSGE-VAR whereas the habit persistence parameter and to a lesser extent the intertemporal elasticity of substitution turns out somewhat higher. Moreover, the Ramsey estimation delivers a higher degree of price indexation than in the Taylor rule DSGE-VAR while the degree of nominal rigidities is lower for prices and wages. Note that the wage indexation parameter in both DSGE-VARs as well as the *Calvo* parameter on wage setting in the Ramsey DSGE-VAR are weakly identified. Finally, the steady state markup in the goods market is slightly lower in the Ramsey DSGE-VAR. Otherwise, the main asymmetries between the two models concern the stochastic processes of the exogenous disturbances and in particular the persistence parameters. The productivity and public expenditure shocks are much more persistent in the Ramsey DSGE-VAR estimation while the Taylor rule specification leads to higher persistence for the consumer preference shock and lower persistence for the monetary policy shock.

[4]In this regard, the approach followed by Del Negro and Schorfheide is, at least computationally, inefficient. Also, contrary to us, they do not average over different possible values of λ but pick a single value of this parameter, which is not the bayesian way.

Table 2 also presents the posterior distributions for the direct estimation of the Ramsey and the Taylor rule DSGE. The comparison with those sets of parameter estimates sheds some light on how the direct Bayesian estimation procedure tries to deal with model misspecifications. Broadly speaking, the direct estimation results amplify some of the differences highlighted in the DSGE-VAR results concerning the behavioral parameters. We obtain a much lower degree of price and wage rigidity as well as of indexation in price-setting for the Ramsey model. Another significant deviation between both models relates to the intertemporal elasticity of substitution which is estimated to be much higher under the optimal policy. Together with a lower habit parameter and higher investment adjustment costs, this reveals that consumption and investment are estimated to be less sensitive to the nominal interest rate in the Ramsey regime. Regarding the stochastic properties of exogenous processes, the main differences between both models relate to higher persistence for the consumer preference shock and lower persistence for the monetary policy shock in the Taylor rule DSGE. Compared with the DGSE-VAR results, we observe that the estimated standard deviations of the shocks and some persistence parameters are higher in the direct estimations for both models.

To sum up, we find as in Del Negro et al. [2007] that the misspecifications affect strongly the stochastic properties of the structural shocks in the direct estimation. Furthermore the additional restrictions imposed by the optimal policy increase the degree of misspecifications of the Ramsey model and result in wider deviations in behaviorial parameter estimates between the Ramsey and the Taylor rule DSGE, when compared with the DSGE-VARs outcome. We learnt from previous contributions that, given a set of estimated parameters for the DSGE, the Ramsey allocation allows for more fluctuations in real quantities while the variations of inflation and especially nominal wage growth are much more muted than with an estimated rule. Consequently, to capture the volatility of wage and price dynamics through the direct estimation, a tension appears between the fit of the price and wage setting curves and the need to mitigate Ramsey planner preference for inflation stabilization. As the welfare cost of nominal rigidities is positively related to the degree of price and wage fixity, the estimation would tend to lower wage and price contract mean durations. But in return, it increases the slope of the supply curves and the required size of the markup shocks, therefore deteriorating the overall empirical performance. This explains notably why the calvo-wage probability is far below the one obtained in the direct DSGE estimation of the Taylor rule model. The same reasoning applies to the real variables, for which the volatility is matched by limiting their sensitivity to the interest rate.

Those results also illustrate the advantages of the DSGE-VAR methodology. When comparing structural models, it is helpful to develop an estimation approach which can account for model misspecifications. In the case at hand, we see that by allowing to relax some of the supplementary cross-restrictions from the optimal policy setting, the structural inference provided by the DSGE-VAR portrays interesting similarities between the Ramsey and the Taylor rule model estimates. The differences observed with the direct estimations reflect the tight policy objectives of the Ramsey allocation, conditioned by the structural parameters and the modeled market imperfections. At the same time, one should keep in mind that, in the polar case $\lambda_{DSGE} = 0$, the DSGE-VAR likelihood is uninformative about the structural parameters. The more the restrictions from the DSGE are relaxed, the less informative is the DSGE-VAR likelihood about the structural parameters.

We now turn to three dimensions of policy evaluation that can be explored from the two estimated DSGE-VARs.

4 Assessing optimality #1: fit of the Ramsey model

A first approach to evaluate the historical performance of monetary policy consists in assessing whether a structural model featuring an optimal policy conduct portrays in a satisfactory manner the statistical properties of the data generating process.

Structural inference based on full information methods have been popular over the recent years and provide likelihood-based criteria to gauge the empirical fit of a DSGE. A blunt test for optimality of historical policy conduct can accordingly be derived from the direct estimation of the Ramsey model. The estimation approach forms a natural basis to construct statistical tests or measures for optimal policymaking. A growing literature has investigated the empirical fit of DSGE models conditional on optimal monetary policy. Among others, Salemi [2006], Dennis [2006] or Favero and Rovelli [2003] estimate such models using full information econometric methods. However, compared with our Ramsey formulation for optimal policy, the authors assume that monetary policy minimizes an *ad hoc* loss function whose relative weights are estimated. Conversely, the Ramsey policy implicitly uses a loss function derived from the quadratic expansion of the aggregate welfare and its weights are linked to the structural parameters of the model.

The estimation method that we promote in this paper is the DSGE-VAR procedure applied to the Ramsey policy model. The results of this first exercise are then systematically put into perspective by comparing with the DSGE-VAR estimation of a Taylor rule model.

One may wonder why we preferred the DSGE-VAR approach to the direct estimation of the Ramsey model which could have been directly compared with the influential results of Smets and Wouters [2007]. The authors show that DSGE models using a Taylor rule can successfully compare with VARs in terms of empirical performance. A fundamental reason for that is linked to Del Negro et al. [2007] which clearly point to non-negligible misspecifications in the modeling framework of Smets and Wouters. Therefore, since the Ramsey policy is likely to introduce tighter restrictions than the Taylor rule specification, a methodology which could control for misspecifications seemed much more appealing to form a judgement on policy comparison. Beyond this, the DSGE-VAR approach also makes use of an explicit reference model and allow to investigates further the modeling dimensions that are not supported by the data (we will come back to that later).

Building on the DSGE-VAR estimations that we described in the previous section, two questions can be raised. First, how good is the Ramsey model in mapping the US data? Second, how does the Ramsey model compare with the Taylor rule specification in terms of empirical performance?

The DSGE-VAR estimation of the structural model with optimal monetary policy provides a first indication of the degree of misspecification of the model. In principle, as soon as the posterior estimates of λ_{DSGE} is different from infinity, it means that the DSGE-VAR empirical performance would be improved by relaxing the restrictions imposed by the structural model on the VAR representation. And in this respect, the posterior mode value for λ_{DSGE} in the Ramsey DSGE-VAR estimation only reaches 1.30 with a 80% highest density interval ranging from 1.02 to 1.70 (see Table 2). Moreover, the log-marginal likelihood of the model is -398.1 which is around 76 points higher than the one obtain with the DSGE ($\lambda_{DSGE} = \infty$). Therefore, significant misspecifications seem present in the Ramsey model which casts some doubts about the ability of the optimal policy to portray appropriately the historical policy conduct. At the same time, the DSGE-VAR estimation delivers a DSGE prior weight λ_{DSGE} which is much higher than the minimum value needed for the prior to be defined, at 0.35 (see appendix B). This suggests that the Ramsey model is nonetheless providing useful prior information for the BVAR.

Comparing now with the Taylor rule specification, the difference in log-marginal data density between the Taylor and the Ramsey DSGE-VARs is around 4.7 which translates into posterior odds of almost 140 to 1 in favor of the Taylor rule DSGE-VAR. The DSGE prior weight is also higher with a posterior mode estimate at 1.57 and a 80% highest density interval ranging from 1.16 to 2.18. The likelihood comparison presented here applies to the DSGE-VAR models and not to the DSGEs. The posterior odds ratio therefore only tells us that the Taylor rule specification is preferred as prior structure for a BVAR. In the DSGE approach, ignoring the misspecification problem makes the data even less supportive for the optimal policy. The marginal likelihoods discrepancy amounts to around 32.2.

All in all, likelihood-based measures explored in this section point to significant *distance* between the Ramsey model and either an agnostic VAR or a Taylor rule specification. In particular, if one concludes from this exercise that the estimated Taylor rule is the best representation of monetary policy conduct, the hypothesis that historical monetary policy conduct has been optimal would then be soundly rejected.

Table 3: **Moments and RMSE from the structural model:** comparison between the DSGE-VAR and the DSGE estimations.

	DSGE-VAR		**DSGE** direct		**Data**
	Taylor	Ramsey	Taylor	Ramsey	84Q1-07Q3
Standard deviation					
ΔY_t	0.52	0.50	0.67	0.65	0.52
ΔC_t	0.39	0.38	0.59	0.58	0.49
ΔI_t	1.22	1.17	1.83	1.75	1.67
L_t	0.82	0.77	3.63	2.80	1.83
Δw_t	0.60	0.60	0.80	0.80	0.78
Π_t	0.20	0.19	0.50	0.43	0.24
R_t	0.22	0.18	0.47	0.35	0.57
Correlations					
ΔY_t, ΔC_t	0.64	0.65	0.62	0.60	0.53
ΔY_t, ΔI_t	0.58	0.51	0.61	0.57	0.59
ΔY_t, Δw_t	0.16	0.20	0.21	0.23	0.19
ΔY_t, Π_t	-0.16	-0.19	-0.25	-0.22	-0.19
RMSE *in sample*					BVAR(4)
ΔY_t	0.47	0.68	0.49	0.52	0.54
ΔC_t	0.52	0.74	0.52	0.52	0.50
ΔI_t	1.75	1.79	1.43	1.48	1.29
L_t	0.87	1.02	0.86	0.71	0.35
Δw_t	0.79	1.00	0.78	0.78	0.70
Π_t	0.27	0.29	0.21	0.23	0.18
R_t	0.16	0.19	0.13	0.13	0.11

At the same time, it is crucial to acknowledge that model comparison in general, and selection of specific behavioral structures in particular, based on marginal density can fail to provide satisfying robustness, as pointed out by Sims [2003]. In order to put into perspective the difference of log-marginal data density that we find between the Ramsey and the Taylor rule DSGE-VAR, we estimated a DSGE-VAR with a slightly different specification than in Smets and Wouters [2007]. Removing the term on output gap first difference in the policy rule deteriorates the performance of the DSGE-VAR (results not reported here), leading to a log-marginal data density of -400.2 which is even lower than the one obtain with the Ramsey policy. The DSGE prior weight is nonetheless higher than in the Ramsey DSGE-VAR at 1.42 for the posterior mode estimate.

Even if the log marginal likelihood comparisons clearly favor the Taylor model over the optimal policy model, it is also important to investigate where the rejected structural model fails. Beyond the comparison of marginal density, we thus examine the relative performance of the Taylor rule and Ramsey DSGEs in terms of in sample RMSEs and second order moments. The marginal likelihood capturing the relative one-step-ahead predictive performance of a model, the in sample one-quarter-ahead RMSEs can help us to gain intuition on what drives the reported posterior odds analysis. Table 3 presents RMSEs, unconditional standard deviations and main correlations in the data and implied by the Ramsey and Taylor rule models, evaluated either at the posterior mode from the DSGE-VAR estimation or from the direct DSGE estimates. We also report the RMSEs of a four lags Bayesian VAR estimated with Minnesota-type prior. The RMSEs appears to be quite close across the monetary policy regimes with the direct estimation parameters. However, when using the DSGE-VAR estimates, the Taylor rule model generates lower RMSEs for almost all variables, with more moderate gains on inflation and interest.

Turning to the second order moments, overall, the volatilities are slightly lower in the Ramsey model than in the Taylor rule model, for both sets of parameter estimates. The standard deviations of interest rate in the Ramsey DSGE is sensibly lower than in the Taylor rule DSGE which reflects partly the relatively high welfare penalty for instrument fluctuations λ_r. Compared with sample moments, the standard deviations are lower with the DSGE-VAR parameter estimates and higher with the DSGE parameter estimates. Studies analyzing optimal policy within estimated medium-scale DSGE models like Adjemian et al. [2007] or Adjemian et al. [2008] indicate that the Ramsey allocation is likely to induce significantly lower volatility of inflation and higher volatility of real variables than under estimated Taylor rule specifications. But, we see that, when bringing the Ramsey model to the data, it can somewhat match the main moments and correlations qualitatively as well as the Taylor rule model.

Overall, the analysis of the empirical performance for the Ramsey model confirms the *ex ante* intuition that the welfare-maximizing monetary policy does not provide enough degree of freedom to match US data, compared with a Taylor rule specification. At the same, the fit along selected dimension, either through the DSGE-VAR or for moments, surprised us positively.

5 Assessing optimality #2: welfare cost of the Taylor rule

A second approach of policy evaluation that has been popular in the literature concerns the use of welfare analysis to assess the properties of estimated policy rules. Among others, Adjemian et al. [2007] on the euro area and Levin et al. [2005] on the US, estimate DSGE models based on Taylor rule specifications and then use the behavioral parameters to analyze optimal monetary policy and the welfare costs of alternative rules.

Along this dimension, the analytical framework presented in this paper offers sensible contributions. First, given that the DSGE-VAR estimation seems to support the Taylor rule specification, it would be consistent to take the posterior parameter distribution of this model and then assess the *distance* to optimality of historical monetary policy conduct through welfare cost measures with respect to the Ramsey allocation. Second, we illustrate further the difference between the Taylor and the optimal allocation by computing welfare-based simple optimal rules.

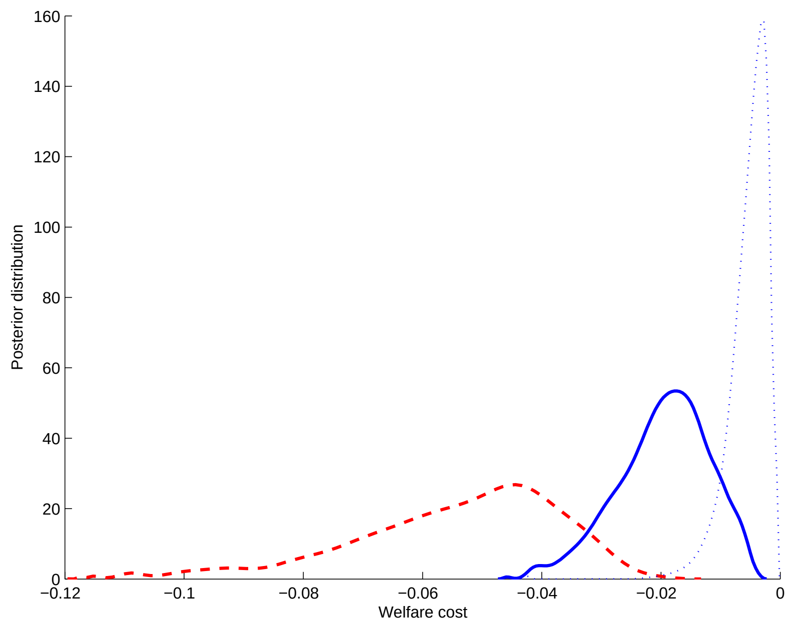
The welfare costs comparisons are performed using welfare measures conditional on the steady state Ramsey allocation. More specifically, we compute the fraction of consumption stream from alternative monetary policy regime to be added (or subtracted) to achieve the reference level corresponding to the allocation following the estimated policy rule. The welfare cost, in percentage points, is then given by $welfarecost = \psi \times 100$ with

$$\psi = \left[\frac{\mathcal{W}_t^{est}}{\mathcal{W}_t^{Ramsey}}\right]^{\frac{1}{1-\sigma_c}} - 1$$

where $\mathcal{W}_t^{est}$ is the welfare obtained under the estimated policy rule and $\mathcal{W}_t^{Ramsey}$ the one under the optimal policy regime.

Figure 1 presents the welfare cost distribution of the estimated Taylor rule using parameter uncertainty derived from the DSGE-VAR estimation of the Taylor model. The policy parameters (the coefficient of the Taylor rule and the penalty on interest rate fluctuations for the Ramsey policy) are kept constant at the posterior mode of their respective estimation and we remove the interest rate shock. The welfare cost of the estimated Taylor rule amounts to 0.017% of steady state consumption at the mode, when using the Taylor DSGE-VAR parameter distribution. Such welfare calculations could also be put into perspective by doing the same exer-

Figure 1: Welfare costs of the estimated rule and the optimized rule on the Volker-Greenspan sample .



Note: posterior parameter distributions from the Taylor rule DSGE-VAR estimation (plain line = estimated rule; dotted line = optimized rule) and from the direct estimation of the Taylor rule DSGE (dashed line).

cise with the posterior parameter distributions from the Ramsey DSGE-VAR estimation which would imply a lower welfare cost of around 0.005% (distribution not reported here).

Those levels contrast with higher absolute values, averaging 2%, reported for example by Adjemian et al. [2007] for the euro area, albeit with different utility specification, shock structure and estimation methodology. The estimated Taylor rule thus apparently turns out to perform relatively well from a welfare perspective. Nonetheless, the optimal policy literature clearly indicates that welfare assessment and policy ranking may not be robust to alternative shock structures, steady state inefficiency or real rigidities. In particular, the low absolute welfare cost we obtain, could be due to the fact that the DSGE-VAR estimation may have reduced autocorrelations and standard deviations of exogenous disturbances compared with a direct estimation (see Del Negro et al. [2007] for a related point). A direct estimation of the Taylor rule model would actually deliver a conditional welfare cost more than twice higher at 0.045%, evaluated at the posterior mode of the parameters. In order to explore the sources of this gap, we restricted persistence parameters and shock standard deviations to be the same as in the Taylor rule DSGE-VAR estimation. The welfare cost then shrinks to 0.015%.

A second perspective on the welfare cost of the estimated Taylor rule comes from the comparison with welfare-based optimal rules. Given the Taylor DSGE-VAR structural parameters, we computed the interest rate rule, based on the same target variables as in the estimated rule, which maximizes the aggregate welfare augmented with the penalty for interest rate fluctuations. In contrast to the estimated Taylor rule, we allowed for an AR(2) term in the interest rate rule. We obtain the following optimal coefficients:

$$\begin{array}{rcl} \text{Optimal} & : & \hat{R}_t = \underset{2.662}{\rho_1} \hat{R}_{t-1} - \underset{1.523}{\rho_2} \hat{R}_{t-2} + \underset{0.148}{\widetilde{r}_\pi} \hat{\pi}_{t-1} + \underset{0.043}{\widetilde{r}_y} \hat{y}_{t-1} + \underset{0.544}{\widetilde{r}_{\Delta y}} \Delta\hat{y}_t \\ \text{Estimated} & : & \hat{R}_t = \underset{0.836}{\rho_1} \hat{R}_{t-1} + \underset{0.263}{\widetilde{r}_\pi} \hat{\pi}_{t-1} + \underset{0.026}{\widetilde{r}_y} \hat{y}_{t-1} + \underset{0.026}{\widetilde{r}_{\Delta y}} \Delta\hat{y}_t \end{array}$$

The optimal rule is characterized by standard features emphasized in the theoretical literature on optimal policy (see for example Giannoni and Woodford [2003]). First, we find as expected a *super inertia* on interest rates, which guided our AR(2) specification: the optimal rule implies not only intrinsic inertia in the dynamics of the interest rate (since a transitory deviation of the inflation rate from its average value increases the interest rate in both the current quarter and the subsequent quarter), but also induces an explosive dynamic for the interest rate if the initial overshooting of the long-run average inflation rate is not offset by a subsequent undershooting (which actually always happens in equilibrium). Second, the difference term on the output gap enters the rule with a much higher coefficient than for the level term which is consistent with optimal targeting rules derived within much simpler setups (see Woodford [2003]). Compared with the estimated rule, the optimal one puts more weight on the model-based output gap stabilization and less on the inflation.

The welfare cost implied by the optimal rule is reduced to around 0.004%. This remaining cost highlights the intrinsic sub-optimality of the estimated Taylor rule due to its specification. As shown in Adjemian et al. [2007] for a similar exercise, adding wage inflation in the optimal rule delivers a higher welfare and is consistent with theoretical rules implementing the Ramsey allocation in simpler modeling frameworks.

Overall, while the deterioration in welfare associated with the estimated Taylor rule seems quantitatively modest, we refrain from drawing strong conclusions. Instead, we would like to emphasize that welfare-based policy evaluation remains quite sensitive to model dimensions that cannot be easily captured through statistical inference of the first order approximation of the model, based on macroeconomic data.

6 Assessing optimality #3: conditional propagation

A third dimension of policy evaluation which has been explored within the optimal monetary policy literature arises from limited information approaches. Gali et al. [2003] and Avouyi-Dovi and Matheron [2007] among others, have relied on SVAR evidence about the macroeconomic transmission of technological shocks, to which DSGE models embedding optimal policy setting have been confronted. Partial information inference based on minimum distance techniques advocated by Christiano et al. [2005] for example, allow then to construct formal statistical test about the optimality of historical monetary policy conduct.

The comparison between the structural model and the SVAR impulse responses bears some crucial limitations. Obviously, as an empirical benchmark, the VAR should provide a better statistical performance than the structural model. But recent work has shown that unrestricted VAR does not improve on the fit and the forecasting ability of medium-scale DSGE. Moreover, the identification of structural shocks in the context of a VAR requires auxiliary assumptions which have to be model-consistent so that, if the DSGE is the reality, impulse responses should coincide.

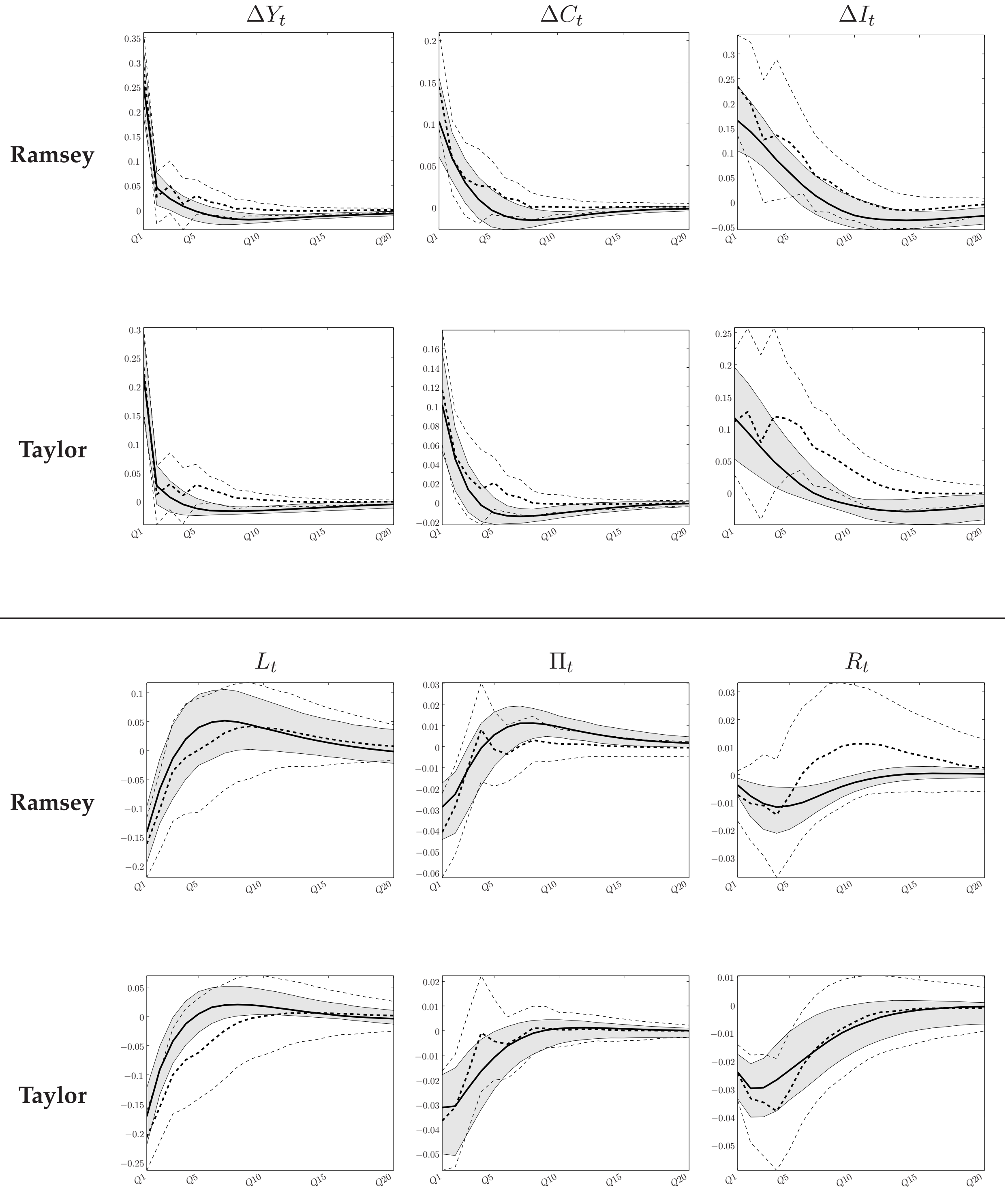
One major contribution of the DSGE-VAR methodology from Del Negro and Schorfheide [2004] was to address those pitfalls in a consistent manner. The estimated DSGE-VAR constitutes a useful benchmark model satisfying the requirements of empirical performance and model-consistency of the identification scheme. And in this paper, this approach enables us to pursue a policy evaluation along various conditional dimensions and therefore extending the previous literature to a wider set of structural disturbances.

The estimation of both the Ramsey and the Taylor DSGE-VARs provides various degrees of comparison to assess the *optimality* of historical monetary response to selected structural disturbances. First of all, even if the Ramsey model suffers from significant misspecifications as highlighted by the DSGE-VAR estimation, the methodology also provides a reference structural VAR which allows to compare impulse response functions (IRFs). And the Ramsey model could indeed perform well in the macroeconomic transmission of certain shocks. A second degree of analysis can be drawn from the comparison with the DSGE-VAR and DSGE propagation mechanism based on the Taylor rule specification: assuming that some IRFs are very close in the Ramsey DSGE and its corresponding SVAR, it is interesting to see whether the same holds within the Taylor DSGE-VAR framework. Finally, given that the identification scheme used in the DSGE-VAR approach is model dependent, one may examine the possible differences in the economic transmission implied by the Ramsey or the Taylor rule models even in the cases where DSGE and DSGE-VAR IRFs are similar for both monetary policy specifications.

We concentrate first on the impulse responses of a technological shock (see Figures in table 4), not least because this disturbance has attracted a lot of attention in the related literature but also due to the striking similarities within and across our two DSGE-VAR setups.

On average, the Ramsey policy does well in this dimension by all means of comparison: the model-based IRFs are very close to the ones of the structural VAR; the small distance between DSGE and DSGE-VAR IRFs with the Ramsey policy are comparable to what is obtained with the Taylor rule; and the broad economic transmission is qualitatively analogous with both policy regimes. Such results support the findings of studies based on the minimization of IRFs between VAR evidence and structural models which concluded that monetary policy behaved optimally in response of technological shock (see for example Avouyi-Dovi and Matheron [2007]). This result is particularly worth emphasizing in our case since the estimation methodology was not directed to match the VAR responses of any specific shock. At the margin, some differences emerge: in the Ramsey allocation, the negative inflationary pressures stemming from a positive productivity shock are less persistent than in the DSGE-VAR and in the Taylor rule IRFs, with inflation even bouncing back to positive territory after one year and a half. The interest rate path also presents some discrepancies: compared with the Taylor model, the Ramsey policy leads to a smaller decline in the policy rate.

Table 4: Transmission of a productivity shock in the DSGE-VAR *(dotted lines)* and in the DSGE *(plain lines)*: *Ramsey and Taylor DSGE-VARs on the Volker-Greenspan sample.*



The results for the other structural shocks are presented in Tables 6 and 7 in the appendix. We see first that most of the previous comments extend to the wage markup shock. IRFs in the DSGE and in the DSGE-VAR are relatively similar for the Ramsey model on the one hand and for the Taylor rule model on the other hand. One exception is the lower persistence of the inflation response in the Ramsey allocation which is not supported neither by the its corresponding DSGE-VAR nor by the Taylor rule model IRFs. However, even if the distance between the DSGE and the DSGE-VAR responses for each policy formulation are broadly analogous, the Ramsey IRFs feature a stronger adjustment in real quantities and more moderate disinflationary effects.

Discrepancies increases for the others shocks. Admittedly, the responses of GDP, consumption, investment and real wages to an investment shock or a government spending shock are relatively similar under the Ramsey policy and the estimated rule. However the inflation response is slightly negative in the Ramsey allocation contrary to its DSGE-VAR counterpart and to the IRFs based on the Taylor rule model. The transmission of a preference shock also reveals a high degree of misspecification in the Ramsey model while the IRFs for this shock are relatively similar in the Taylor rule DSGE and DSGE-VAR. The initial increase in real quantities is too short-lived in the Ramsey model and the response of hours is particularly weak compared to the DSGE-VAR and its Taylor rule counterparts. Most importantly, the Ramsey policy features almost an opposite response of inflation to what its DSGE-VAR and the Taylor rule model would suggest. Under the estimated rule, the preference shock is expansionary on GDP and upward pressures emerge on inflation.

Considering price-markup shocks, the transmission to the economy as shown by the DSGE-VARs benchmarks is not very well captured in any of the two estimated models. Moreover, apart from the interest rate, the two DSGE-VARs exhibit almost the same dynamics and the same deviations to the DSGEs. The Ramsey thus obviously inherits the misspecifications of the Taylor model and worsens the picture on the interest rate.

A final comment relates to the interest rate shock. With the Taylor rule specification, this shock is interpreted as a non-systematic monetary policy impulse and presents a very similar transmission in the DSGE and in the DSGE-VAR. For the Ramsey policy, the way it has been introduced makes it very similar to a negative preference shocks, as we already mentioned, except for the interest rate dynamics. That is why the positive interest rate shock dampens real variables but implies a positive inflation response. It is obvious that from an optimal policy perspective there is a weak economic rationale in introducing Taylor rule residuals and that other source of volatility would be more appropriate. We did not investigate alternative shock structure for the Ramsey allocation in the present paper in order to keep the symmetry with the well-established Smets and Wouters specification.

7 Results from the pre-Volker sample

In this final section, we revisit the results obtained previously by estimating the Taylor and Ramsey DSGE-VARs on the *Great Inflation* period. The corresponding sample ranges from 1966Q1 to 1979Q3 and ends with the appointment of Paul Volker as Chairman of the Federal Reserve Board.

The posterior distribution of parameter estimates are presented in Table 5. One difference with respect to the Volker-Greenspan sample results concerns the nominal rigidity coefficients. Both for the Ramsey and the Taylor rule models, the degree of price and wage stickiness decreases while the degree of indexation in price setting is higher (the indexation parameter on the wage setting is here again not identified). This result is consistent with the findings of Smets and Wouters [2007] and the widely-shared view that over the recent decades, the Phillips curve has flattened and become less backward looking.

Regarding the relative empirical performance of the Ramsey model compared with the Taylor model over the Great Inflation period, we find a difference of 6 points of log-marginal data density which is slightly higher than what we obtain on the most recent sample, but there is not compelling evidence that the Ramsey model does worse on the pre-Volker sample. The welfare cost analysis also points slightly to the same conclusion suggesting that the *distance to optimality* of the estimated Taylor rule over this period was similar to the one obtained over the recent decades. This result may be due to an important caveat which applies to our policy evaluation over this period. The estimation of the Taylor rule model was conducted ruling out, by assumption, the possibility that U.S. monetary policy during the 70's had been significantly worse than it has been over the most recent period since the parameters space was restricted to the determinacy region. However, the possibility that, before October 1979, U.S. monetary policy had been so weakly counter-inflationary as to put the economy in the alternative indeterminacy region - characterized by an intrinsically larger macroeconomic volatility across the board - is at the source of the *bad policy* interpretation of the Great Inflation as exposed by Clarida et al. [2000].

Finally, regarding the impulse responses presented of Tables 8 and 9 in the appendix, the most observations made previously also hold for the pre-Volker estimates. One striking difference however concerns the transmission of the price markup shock in the Taylor rule model and to a lesser extent in the Ramsey model, which becomes much closer to the DSGE-VAR propagation.

8 Concluding remarks

Overall, the present paper intends to bring a methodological contribution to the abundant literature on monetary policy evaluation. Through the DSGE-VAR estimation of medium-scale DSGE model featuring welfare-maximizing monetary policy and the comparison with a benchmark Taylor rule specification, we propose a unified framework to cover the main approaches which *provocatively* assess the *optimality* of historical monetary policy conduct.

Using US data over the Volker-Greenspan, our results suggest that the Taylor rule specification provides a better description of US data than the Ramsey model. At the same time, while the statistical inference supports the Taylor rule model, counterfactual analysis points to relatively modest welfare costs of such a policy compared with the optimal allocation. Finally, the comparison of impulse response functions in the DSGE-VAR and the DSGE for the Ramsey and the Taylor rule models shows that the transmission of a productivity shock to the US economy is very similar across all dimensions. However, this conclusion does not hold for other type of disturbances like consumer preference shocks for example.

References

S. Adjemian, M. Darracq Pariès, and S. Moyen. Optimal monetary policy in an estimated dsge for the euro area. Working Paper 803, European Central Bank, 2007.

S. Adjemian, M. Darracq Pariès, and S. Smets. A quantitative perspective on optimal monetary policy cooperation between the us and the euro area. Working Paper 884, European Central Bank, 2008.

Sanvi Avouyi-Dovi and Julien Matheron. Technology shocks and monetary policy: Revisiting the fed's performance. *Journal of Money, Credit and Banking*, 39(2-3):471–507, 2007.

L. Christiano, M. Eichenbaum, and C. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45, 2005.

R. Clarida, J. Gali, and M. Gertler. Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics*, (1):147–180, 2000.

M. Del Negro and F. Schorfheide. Priors from general equilibrium models for vars. *International Economic Review*, 45:643–673, 2004.

M. Del Negro, F. Schorfheide, F. Smets, and R. Wouters. On the fit of new-keynesian models. *Journal of Business and Economic Statistics*, 25(2):123–143, 2007.

Richard Dennis. The policy preferences of the us federal reserve. *Journal of Applied Econometrics*, 21(1):55–77, 2006.

A. Dixit and J. Stiglitz. Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308, 1977.

Michael Dotsey and Robert G. King. Implications of state-dependent pricing for dynamic macroeconomic models. *Journal of Monetary Economics*, 52(1):213–242, 2005.

Carlo A Favero and Riccardo Rovelli. Macroeconomic stability and the preferences of the fed: A formal analysis, 1961-98. *Journal of Money, Credit and Banking*, 35(4):545–56, 2003.

Jordi Gali, J. David Lopez-Salido, and Javier Valles. Technology shocks and monetary policy: assessing the fed's performance. *Journal of Monetary Economics*, 50(4):723–743, 2003.

M. Giannoni and M. Woodford. Optimal interest-rate rules: II. Applications. Working Paper 9420, NBER, 2003.

A. Khan, R. King, and A. Wolman. Optimal monetary policy. *Review of Economic Studies*, 70(4): 825–860, 2003.

M. Kimball. The quantitative analysis of the basic neomonetarist model. *Journal of Money, Credit and Banking*, 27(4):1241–1277, 1995.

A. Levin, A. Onatski, J. Williams, and N. Williams. Monetary policy under uncertainty in micro-founded macroeconomic models. Working Paper 11523, NBER, 2005.

J. Rotemberg and M. Woodford. An otpimization-based econometric framework for the evaluation of monetary policy. *NBER Macroecnomics Annual*, pages 297–344, 1997.

M. Salemi. Econometric policy evaluation and inverse control. *Journal of Money, Credit and Banking*, 38(7):1737–1764, 2006.

S. Schmitt-Grohe and S. Uribe. Optimal inflation stabilization in a medium-scale macroeconomic model. Working Paper 11854, NBER, 2005.

C. Sims. Probability models for monetary policy decisions. Manuscript, Princeton University, 2003.

F. Smets and R. Wouters. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123–1175, 2003.

F. Smets and R. Wouters. Comparing shocks and frictions in us and euro area business cycles: a bayesian dsge approach. *Journal of Applied Econometrics*, 20(1), 2005.

Frank Smets and Rafael Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606, 2007.

M. Woodford. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, 2003.

A Recursive formulation of price and wage settings

A.1 Wage setting

In the following, given that the steady state model features a balanced growth path, all variables are appropriately deflated to be stationary in the stochastic equilibrium.
The first order condition of the union's program for the re-optimized wage w_t^* can be written recursively as follows:

$$w_t^* = \frac{\theta_w(1+\psi_w)}{(\theta_w(1+\psi_w)-1)} \frac{\mathcal{H}_{1,t}}{\mathcal{H}_{2,t}} + \frac{\psi_w}{(\theta_w-1)} (w_t^*)^{1+\theta_w(1+\psi_w)} \frac{\mathcal{H}_{3,t}}{\mathcal{H}_{2,t}} \tag{18}$$

with

$$\begin{aligned} \mathcal{H}_{1,t} &= \varepsilon_t^B \widetilde{L} L_t^{1+\sigma_l} w_t^{\theta_w(1+\psi_w)} \left(C_t - \eta C_{t-1}/\gamma\right)^{(1-\sigma_c)} \exp^{\left(\widetilde{L}\frac{(\sigma_c-1)}{(1+\sigma_l)} L_t^{(1+\sigma_l)}\right)} \Delta_{w\lambda,t}^{\theta_w(1+\psi_w)/(1-\theta_w(1+\psi_w))} \\ &\quad + \beta\gamma^{(1-\sigma_c)}\alpha_w \mathbb{E}_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\xi_w}\left[\pi^\star\right]^{(1-\xi_w)}} \right)^{\theta_w(1+\psi_w)} \mathcal{H}_{1,t+1} \right] \end{aligned} \tag{19}$$

$$\begin{aligned} \mathcal{H}_{2,t} &= (1-\tau_{w,t})\lambda_t L_t w_t^{\theta_w(1+\psi_w)} \Delta_{w\lambda,t}^{\theta_w(1+\psi_w)/(1-\theta_w(1+\psi_w))} \\ &\quad + \beta\gamma^{(1-\sigma_c)}\alpha_w \mathbb{E}_t \left[\left(\frac{\pi_{t+1}}{\pi_t^{\xi_w}\left[\pi^\star\right]^{(1-\xi_w)}} \right)^{\theta_w(1+\psi_w)-1} \mathcal{H}_{2,t+1} \right] \end{aligned} \tag{20}$$

$$\mathcal{H}_{3,t} = (1-\tau_{w,t})\lambda_t L_t + \beta\gamma^{(1-\sigma_c)}\alpha_w \mathbb{E}_t \left[\left(\frac{\pi_t^{\xi_w}\left[\pi^\star\right]^{(1-\xi_w)}}{\pi_{t+1}} \right) \mathcal{H}_{3,t+1} \right] \tag{21}$$

The aggregate wage dynamics could also be expressed as

$$\begin{aligned} (w_t)^{1-\theta_w(1+\psi_w)} \Delta_{w\lambda,t} &= (1-\alpha_w) (w_t^*)^{1-\theta_w(1+\psi_w)} \\ &\quad + \alpha_w \left(\frac{\pi_t}{\pi_{t-1}^{\xi_w}\left[\pi^\star\right]^{1-\xi_w}} \right)^{\theta_w(1+\psi_w)-1} (w_{t-1})^{1-\theta_w(1+\psi_w)} \Delta_{w\lambda,t-1} \end{aligned} \tag{22}$$

The previous equations include a dispersion index $\Delta_{w\lambda,t}$ which is related to the re-optimizing wage and the aggregate wage through the following conditions

$$1 = \frac{1}{1+\psi_w} \Delta_{w\lambda,t}^{1/(1-\theta_w(1+\psi_w))} + \frac{\psi_w}{1+\psi_w} \Delta_{wl,t} \tag{23}$$

$$\Delta_{wl,t} = (1-\alpha_w) \left(\frac{w_t^*}{w_t} \right) + \alpha_w \left(\frac{w_t}{w_{t-1}} \frac{\pi_t}{\pi_{t-1}^{\xi_w}\left[\pi^\star\right]^{1-\xi_w}} \right)^{-1} \Delta_{wl,t-1} \tag{24}$$

The market clearing condition linking total labor demand of intermediate firms and total labor supply of households includes a wage dispersion index given by

$$\Delta_{wk,t} = \frac{1}{1+\psi_w} \Delta_{w,t} \cdot \Delta_{w\lambda,t}^{\theta_w(1+\psi_w)/(1-\theta_w(1+\psi_w))} + \frac{\psi_w}{1+\psi_w} \tag{25}$$

with

$$\Delta_{w,t} = (1-\alpha_w)\left(\frac{w_t^*}{w_t}\right)^{-\theta_w(1+\psi_w)} + \alpha_w\left(\frac{w_t}{w_{t-1}}\frac{\pi_t}{\pi_{t-1}^{\xi_w}[\pi^\star]^{1-\xi_w}}\right)^{\theta_w(1+\psi_w)}\Delta_{w,t-1} \tag{26}$$

A.2 Price setting

The first order condition of the intermediate firms profit maximization leads to

$$p_t^* = \frac{\theta_p(1+\psi)}{(\theta_p(1+\psi)-1)}\frac{\mathcal{Z}_{1,t}}{\mathcal{Z}_{2,t}} + \frac{\psi}{(\theta_p-1)}(p_t^*)^{1+\theta_p(1+\psi)}\frac{\mathcal{Z}_{3,t}}{\mathcal{Z}_{2,t}} \tag{27}$$

with

$$\begin{aligned}
\mathcal{Z}_{1,t} &= \lambda_t mc_t Y_t \Delta_{p\lambda,t}^{\theta_p(1+\psi)/(1-\theta_p(1+\psi))} \\
&\quad + \beta\gamma^{(1-\sigma_c)}\alpha_p\mathbb{E}_t\left[\left(\frac{\pi_{t+1}}{\pi_t^{\xi_p}[\pi^\star]^{(1-\xi_p)}}\right)^{\theta_p(1+\psi)}\mathcal{Z}_{1,t+1}\right] && (28)\\
\mathcal{Z}_{2,t} &= (1-\tau_{p,t})\lambda_t Y_t \Delta_{p\lambda,t}^{\theta_p(1+\psi)/(1-\theta_p(1+\psi))} \\
&\quad + \beta\gamma^{(1-\sigma_c)}\alpha_p\mathbb{E}_t\left[\left(\frac{\pi_{t+1}}{\pi_t^{\xi_p}[\pi^\star]^{(1-\xi_p)}}\right)^{\theta_p(1+\psi)-1}\mathcal{Z}_{2,t+1}\right] && (29)\\
\mathcal{Z}_{3,t} &= (1-\tau_{p,t})\lambda_t Y_t + \beta\gamma^{(1-\sigma_c)}\alpha_p\mathbb{E}_t\left[\left(\frac{\pi_t^{\xi_p}[\pi^\star]^{(1-\xi_p)}}{\pi_{t+1}}\right)\mathcal{Z}_{3,t+1}\right] && (30)
\end{aligned}$$

Aggregate price dynamics can then be written as

$$\Delta_{p\lambda,t} = (1-\alpha_p)(p_t^*)^{1-\theta_p(1+\psi)} + \alpha_p\left(\frac{\pi_t}{\pi_{t-1}^{\xi_p}[\pi^\star]^{1-\xi_p}}\right)^{\theta_p(1+\psi)-1}\Delta_{p\lambda,t-1} \tag{31}$$

Here again, compared with the Dixit-Stiglitz aggregator case, the previous equations include a dispersion index $\Delta_{p\lambda,t}$ which is given by

$$\begin{aligned}
1 &= \frac{1}{1+\psi}\Delta_{p\lambda,t}^{1/(1-\theta_p(1+\psi))} + \frac{\psi}{1+\psi}\Delta_{pl,t} && (32)\\
\Delta_{pl,t} &= (1-\alpha_p)(p_t^*) + \alpha_p\left(\frac{\pi_t}{\pi_{t-1}^{\xi_p}[\pi^\star]^{1-\xi_p}}\right)^{-1}\Delta_{pl,t-1} && (33)
\end{aligned}$$

The market clearing conditions in the goods market also involves a price dispersion index given by

$$\Delta_{pk,t} = \frac{1}{1+\psi}\Delta_{p,t}\cdot\Delta_{p\lambda,t}^{\theta_p(1+\psi)/(1-\theta_p(1+\psi))} + \frac{\psi}{1+\psi} \tag{34}$$

with

$$\Delta_{p,t} = (1-\alpha_p)(p_t^*)^{-\theta_p(1+\psi)} + \alpha_p\left(\frac{\pi_t}{\pi_{t-1}^{\xi_p}[\pi^\star]^{1-\xi_p}}\right)^{\theta_p(1+\psi)}\Delta_{p,t-1} \tag{35}$$

B The DSGE-VAR approach

B.1 Deriving the posterior densities

Consider the order p VAR representation for the $1 \times m$ vector of observed variables y_t:

$$y_t = \sum_{k=1}^{p} y_{t-k} \mathbf{A}_k + u_t$$

where $u_t \sim \mathcal{N}(0, \Sigma_u)$. Let z_t be the $mp \times 1$ vector $\left[y'_{t-1}, ..., y'_{t-p}\right]'$ and define $\mathbf{A} = \left[\mathbf{A}'_1, ..., \mathbf{A}'_p\right]'$, the VAR representation can then be written in matrix form as:

$$Y = Z\mathbf{A} + \mathcal{U}$$

where $Y = (y'_1, \ldots, y'_T)'$, $Z = (z'_1, \ldots, z'_T)'$ and $\mathcal{U} = (u'_1, \ldots, u'_T)'$.

Dummy observations prior for the VAR can be constructed using the VAR likelihood function for $\mathcal{T} = [\lambda T]$ artificial data simulated with the DSGE (Y^*, Z^*), combined with diffuse priors. The prior is then given by:

$$p_0 \left(\mathbf{A}, \Sigma \mid Y^*, Z^*\right) \propto |\Sigma|^{-\frac{\lambda T + m + 1}{2}} e^{-\frac{1}{2} tr\left[\Sigma^{-1}\left(Y^{*\prime}Y^* - \mathbf{A}'Z^{*\prime}Y^* - Y^{*\prime}Z^*\mathbf{A} + \mathbf{A}'Z^{*\prime}Z^*\mathbf{A}\right)\right]}$$

implying that Σ follows an inverted Wishart distribution and $\mathbf{A}$ conditional on Σ is gaussian. Assuming that observables are covariance stationary, Del Negro and Schorfheide [2004] use the DSGE theoretical autocovariance matrices for a given $n \times 1$ vector of model parameters θ, denoted $\Gamma_{YY}(\theta)$, $\Gamma_{ZY}(\theta)$, $\Gamma_{YZ}(\theta)$, $\Gamma_{ZZ}(\theta)$ instead of the (artificial) sample moments $Y^{*\prime}Y^*$, $Z^{*\prime}Y^*$, $Y^{*\prime}Z^*$, $Z^{*\prime}Z^*$. In addition, the p-th order VAR approximation of the DSGE provides the first moment of the prior distributions through the population least-square regression:

$$\mathrm{A}^*(\theta) = \Gamma_{ZZ}(\theta)^{-1} \Gamma_{ZY}(\theta) \tag{P1a}$$

$$\Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YZ}(\theta)\Gamma_{ZZ}(\theta)^{-1} \Gamma_{ZY}(\theta) \tag{P1b}$$

Conditional on the deep parameters of the DSGE θ and λ, the priors for the VAR parameters are given by:

$$\begin{aligned} \mathrm{vec}\mathbf{A} \mid \Sigma, \theta, \lambda &\sim \mathcal{N}\left(\mathrm{vec}\mathbf{A}^*(\theta), \Sigma \otimes [\lambda T \Gamma_{ZZ}(\theta)]^{-1}\right) \\ \Sigma \mid \theta, \lambda &\sim \mathcal{IW}\left(\lambda T \Sigma^*(\theta), \lambda T - mp - m\right) \end{aligned} \tag{P2}$$

where $\Gamma_{ZZ}(\theta)$ is assumed to be non singular and $\lambda \geq \frac{mp+m}{T}$ for the priors to be proper[5]. The *a priori* density of $\mathbf{A}$ is defined by $n+1$ parameters (θ and λ), which is likely to be less than mp

[5]Note that it would not be possible to estimate the VAR model by OLS (or maximum likelihood) if we had $\mathcal{T} < m(p+1)$. In this case we would not have more observations than parameters to estimate.

(the VAR number of parameters). If we have a one-to-one relationship (no identification issues) between (θ, λ) and $\mathbf{A}$ it will be a good idea to estimate (θ, λ) instead of $\mathbf{A}$, *ie* to estimate fewer free parameters. To do so, Del Negro and Schorfheide [2004] complete the prior by specifying a prior distribution over the structural model's deep parameters: $p_0(\theta)$. We finally have to set the weight of the structural prior, λ. So we define a prior on the distribution of λ, which is assumed to be independent from θ. Finally, the DSGE-VAR model has the following prior structure:

$$p_0\left(\mathbf{A}, \Sigma, \theta, \lambda\right) = p_0\left(\mathbf{A}, \Sigma \mid \theta, \lambda\right) \times p_0\left(\theta\right) \times p_0\left(\lambda\right) \tag{P3}$$

where $p_0\left(\mathbf{A}, \Sigma \mid \theta, \lambda\right)$ is defined by [P1a,P1b] and [P2].

The posterior distribution, may be factorized in the following way:

$$p\left(\mathbf{A}, \Sigma, \theta, \lambda \mid \mathcal{Y}_T\right) = p\left(\mathbf{A}, \Sigma \mid \mathcal{Y}_T, \theta, \lambda\right) \times p\left(\theta, \lambda \mid \mathcal{Y}_T\right) \tag{Q3}$$

where $\mathcal{Y}_T$ stands for the sample. A closed form expression for the first density function on the right hand side of [Q3] is available. Conditional on θ and λ, [P1a,P1b] and [P2] define a conjugate prior for the VAR model, so its posterior density has to belong to the same family: the distribution of $\mathbf{A}$ conditional on Σ, θ, λ and the sample is matric-variate normal, and the distribution of Σ conditional on θ, λ and the sample is inverted Wishart. More formally, we have:

$$\begin{aligned} \mathrm{vec}\mathbf{A} \mid \Sigma, \theta, \lambda, \mathcal{Y}_T &\sim \mathcal{N}\left(\mathrm{vec}\widetilde{\mathbf{A}}(\theta,\lambda), \Sigma \otimes V(\theta,\lambda)^{-1}\right) \\ \Sigma \mid \theta, \lambda, \mathcal{Y}_T &\sim \mathcal{IW}\left((\lambda+1)T\,\widetilde{\Sigma}(\theta,\lambda), (\lambda+1)T - mp - m\right) \end{aligned} \tag{Q2}$$

where:

$$\widetilde{\mathbf{A}}(\theta,\lambda) = V(\theta,\lambda)^{-1}\left(\lambda T\,\Gamma_{ZY}(\theta) + Z'Y\right) \tag{Q1a}$$

$$\widetilde{\Sigma}(\theta,\lambda) = \frac{1}{(1+\lambda)T}\left[\lambda T\,\Gamma_{YY}(\theta) + Y'Y - \left(\lambda T\,\Gamma_{YZ}(\theta) + Y'Z\right)V(\theta,\lambda)^{-1}\left(\lambda T\,\Gamma_{ZY}(\theta) + Z'Y\right)\right] \tag{Q1b}$$

with:

$$V(\theta,\lambda) = \lambda T\,\Gamma_{ZZ}(\theta) + Z'Z$$

Not surprisingly, we find that the posterior mean of $\mathbf{A}$ is a convex combination of $A^*(\theta)$, the prior mean, and of the OLS estimate of $\mathbf{A}$. When λ goes to infinity the posterior mean shrinks towards the prior mean, *ie* the projection of the DSGE model onto the VAR(p).
We do not have a closed form expression for the joint posterior density of θ and λ (the second term on the right hand side of [Q3]). So the posterior distribution of (θ, λ) is recovered from an MCMC algorithm, as described in [Del Negro and Schorfheide, 2004, appendix B], except that we do estimate λ as the deep parameters θ.[6]

[6]This can be done with Dynare 4.

B.2 Identification

In Del Negro and Schorfheide [2004] the DSGE-VAR approach is shown to provide a quite natural identification scheme for the structural innovations. In the sequel we follow the above mentioned authors. The sole difference is, again related to λ. Our Impulse Response Functions are obtained by averaging over the posterior distribution of λ.

Table 5: **Posterior parameter estimates of Ramsey and Taylor rule DSGE-VARs for the pre-Volker sample.**

	Ramsey				Taylor rule			
	Mode	Mean	$\mathcal{I}_1$	$\mathcal{I}_2$	Mode	Mean	$\mathcal{I}_1$	$\mathcal{I}_2$
σ_c	0.93	0.94	0.67	1.23	1.07	1.09	0.80	1.37
η	0.65	0.63	0.52	0.76	0.65	0.64	0.54	0.75
σ_l	1.31	1.52	0.62	2.47	1.58	1.81	0.82	2.73
ϕ	4.61	4.70	2.75	6.72	4.11	4.24	2.21	6.09
φ	0.52	0.51	0.27	0.76	0.45	0.48	0.24	0.72
α_p	0.48	0.48	0.38	0.58	0.51	0.50	0.38	0.62
ξ_p	0.49	0.49	0.24	0.73	0.33	0.37	0.16	0.58
α_w	0.48	0.48	0.34	0.62	0.59	0.60	0.46	0.72
ξ_w	0.57	0.55	0.33	0.79	0.59	0.57	0.35	0.80
α	0.14	0.14	0.09	0.18	0.15	0.15	0.11	0.19
μ_p	1.22	1.21	1.11	1.34	1.34	1.34	1.19	1.48
r_β	0.25	0.25	0.09	0.40	0.22	0.23	0.07	0.38
γ	0.46	0.45	0.32	0.59	0.43	0.42	0.30	0.55
$\overline{L}$	0.01	0.02	-1.81	1.85	0.31	0.12	-2.04	2.15
$\overline{\pi}$	0.68	0.69	0.50	0.89	0.69	0.69	0.51	0.87
λ_r	0.42	0.46	0.21	0.72	-	-	-	-
r_π	-	-	-	-	1.41	1.45	1.10	1.77
ρ	-	-	-	-	0.78	0.76	0.68	0.85
r_Y	-	-	-	-	0.14	0.14	0.07	0.22
$r_{\Delta Y}$	-	-	-	-	0.20	0.20	0.13	0.27
$\rho_{b,I}$	1.02	1.32	0.41	2.28	0.73	0.97	0.23	1.66
$\rho_{a,g}$	3.20	3.15	1.98	4.48	3.19	3.18	1.87	4.50
ρ_a	0.87	0.82	0.68	0.98	0.95	0.87	0.75	0.99
ρ_b	0.32	0.33	0.11	0.54	0.39	0.38	0.15	0.61
ρ_g	0.79	0.71	0.51	0.94	0.83	0.74	0.53	0.94
ρ_I	0.37	0.38	0.15	0.60	0.47	0.47	0.25	0.70
ρ_p	0.74	0.65	0.38	0.95	0.87	0.74	0.50	0.98
η_p	0.40	0.42	0.14	0.70	0.40	0.43	0.16	0.68
ρ_w	0.88	0.80	0.62	0.97	0.86	0.71	0.43	0.96
η_w	0.60	0.54	0.29	0.79	0.61	0.55	0.28	0.82
ρ_r	0.93	0.86	0.74	0.99	0.29	0.32	0.11	0.52
σ_{ε^a}	0.52	0.53	0.41	0.66	0.51	0.52	0.40	0.63
σ_{ε^b}	1.71	1.83	1.10	2.58	1.69	1.84	1.21	2.47
σ_{ε^g}	2.26	2.28	1.77	2.84	2.42	2.48	1.91	3.01
σ_{ε^I}	4.61	5.00	2.55	7.46	4.25	4.36	2.18	6.45
σ_{ε^p}	0.17	0.17	0.12	0.23	0.14	0.15	0.09	0.20
σ_{ε^w}	0.27	0.29	0.19	0.38	0.17	0.18	0.14	0.22
σ_{ε^r}	0.17	0.17	0.13	0.22	0.25	0.26	0.18	0.33
λ_{DSGE}	2.31	2.67	1.57	3.84	2.54	3.10	1.73	4.56
$P_\lambda(\mathcal{Y})$	-376.52				-368.98			

Table 6: Comparison of impulse responses DSGE-VAR versus DSGE: *Ramsey DSGE-VAR on the Volker-Greenspan sample.*

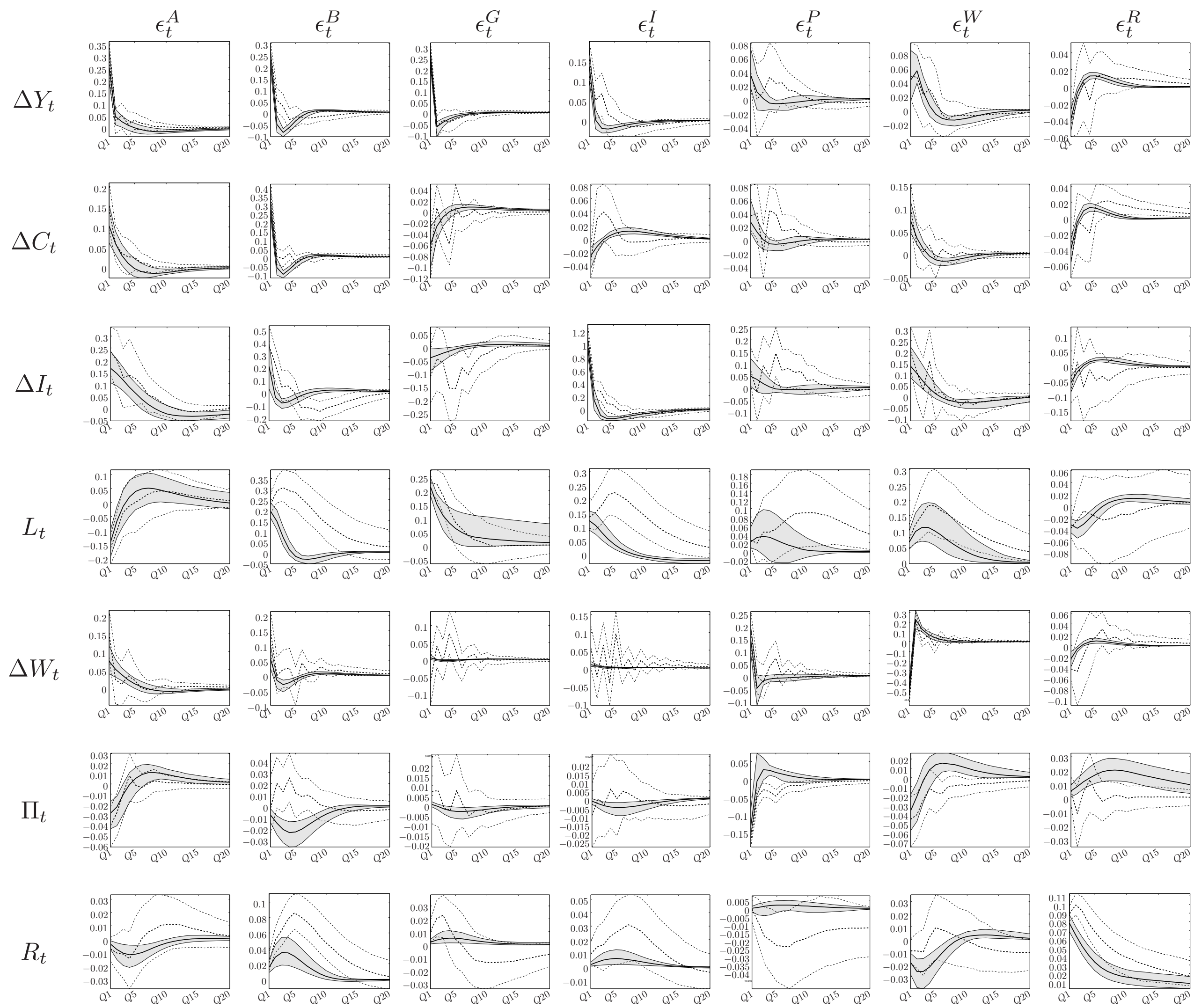


Table 7: Comparison of impulse responses DSGE-VAR versus DSGE: *Taylor rule DSGE-VAR on the Volker-Greenspan sample.*

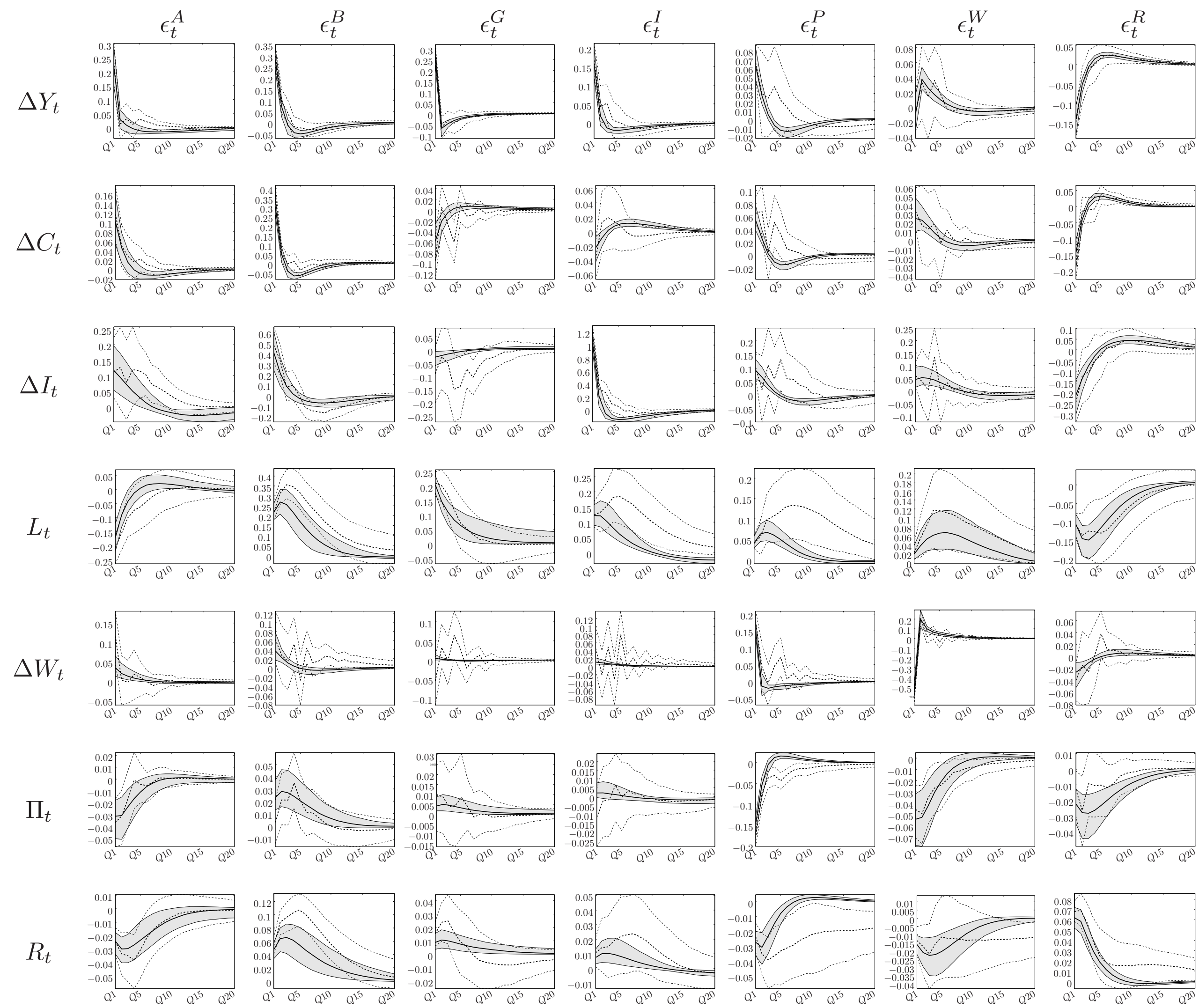


Table 8: Comparison of impulse responses DSGE-VAR versus DSGE: *Ramsey DSGE-VAR on the pre-Volker sample.*

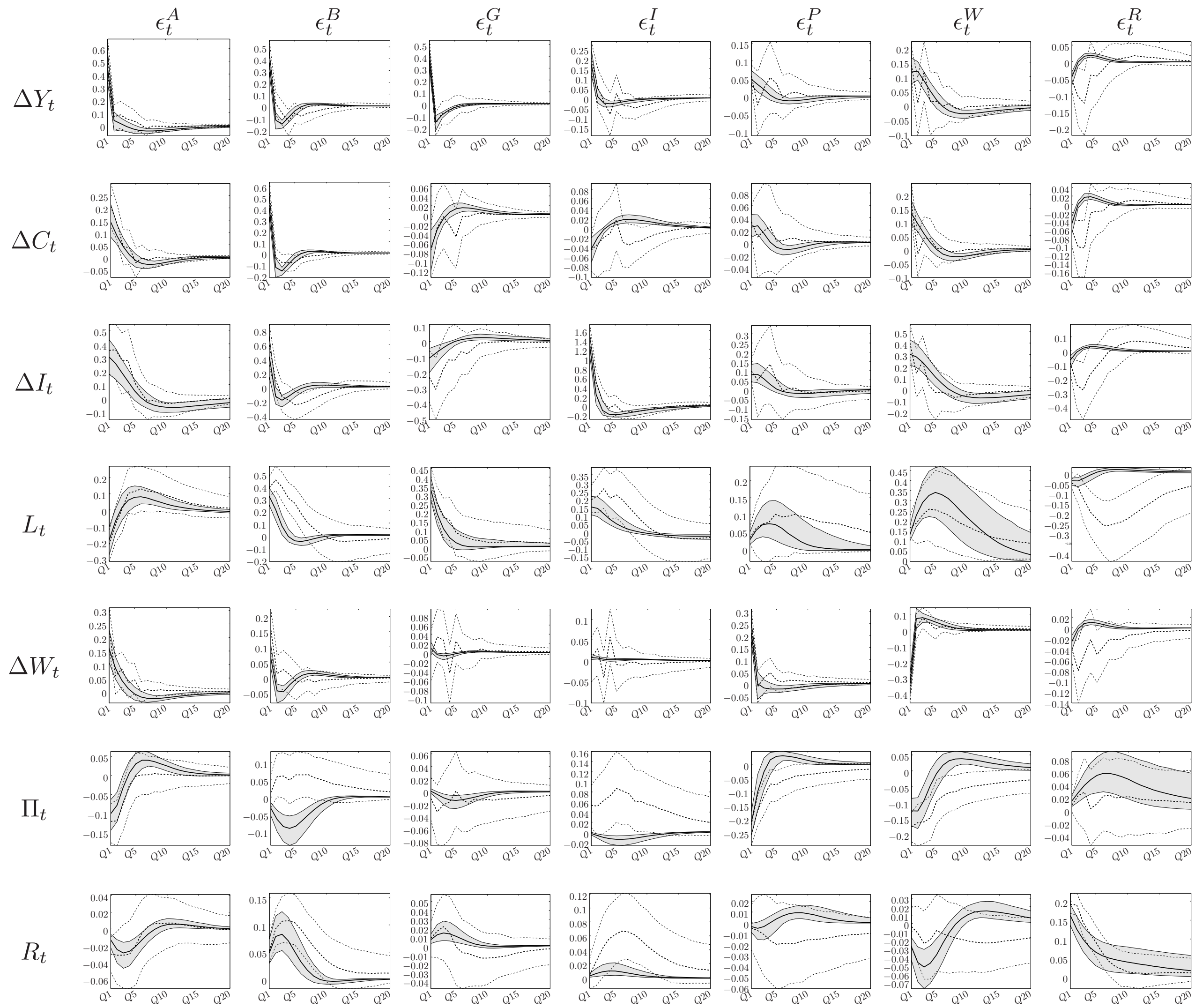


Table 9: Comparison of impulse responses DSGE-VAR versus DSGE: *Taylor rule DSGE-VAR on the pre-Volker sample.*