Measurement with some theory: using sign restrictions to evaluate business cycle models *

Fabio Canova[†]and Matthias Paustian[‡]

November 11, 2007

Preliminary draft

Abstract

We propose a method to evaluate models which is does not require knowledge of the DGP and is robust to the time series specification of the aggregate decision rules. We derive robust (sign) restrictions in a class of models; use some of them to identify structural shocks in the data and others to qualitatively and quantitatively evaluate the model. The approach has good size and excellent power properties; its performance is reasonable even in small samples and the median of the distribution of responses is a good estimator of the true dynamic responses. We use the technique to examine the dynamics of hours in response to technology shocks and of consumption in response to goverment expenditure shocks.

JEL classification: E32, C32.
Keywords: Model validation, Sign restrictions, Structural VARs, Business cycle models.

---

*This paper replaces an earlier paper by the first author which has circulated with the title "Validating Monetary DSGE's with VARs". Canova acknowledges the financial support the Spanish Ministry of Education grant SEJ2004-21682-E (Cambios estructurales and politicas publicas) and from the CREI.

[†]ICREA-UPF, CREI, AMeN, CEPR

[‡]Department of Economics, 302 Business Admin. Bldg, Bowling Green State University, Bowling Green, OH 43403. Phone: (419) 372-3491, Fax: (419) 372-1557, email: *paustim@bgsu.edu*.

1 Introduction

The 1990's have witnessed a remarkable development in the specification of business cycle models: a number of shocks and frictions have been introduced into first generation RBC models driven by a single technological disturbance and our understanding of the propagation mechanism of structural shocks has been considerably enhanced. Steps forward have also been made in estimating such models. While a few years ago it was standard to informally calibrate the structural parameters, now researchers routinely use limited information estimation approaches (see, e.g., Rotemberg and Woodford (1997), Christiano, et. al. (2005)) and successfully implement full information likelihood based procedures (see e.g. Kim (2000), Smets and Wouters (2003), Ireland (2004), Canova (2004), Rabanal and Rubio Ramirez (2005) among many others).

Regardless of the estimation approach one employs, the model used to restrict the data is taken very seriously. In fact, in classical estimation it is implicitly assumed that the model is the DGP of the actual data, up to a set of serially uncorrelated measurement errors, and that the task of the investigator is to find the parameters of this DGP. Since current business cycle models, even in the large scale versions currently used in central banks and international institutions, are still too simple to capture the complexities and the heterogeneities of the macrodata, such an assumption can not be credibly entertained and inference becomes dubious. When a Bayesian framework is employed, the assumption that the model is the DGP is unnecessary to derive the posterior distribution of parameters. Still, even in this framework, it is hard to interpret misspecified estimates, unless an explicit loss function is employed (see Schorfheide (2000)).

Structural estimation faces two additional problems. First, because the aggregate decision rules are non-linear functions of the structural parameters and the mapping is computable only numerically parameter identification becomes an issue (see Canova and Sala (2006)). Second, these methods use considerable computer time and require a dose of ingenuity to solve practical numerical problems.

The 1990s have also witnessed an extraordinary development of structural VAR (SVAR) methods. Structural VARs have enjoyed an increasing success for two reasons: they are easy to estimate and the computational complexities are minimal relative to those of structural techniques; the analysis can be performed without conditioning on a single, and possibly misspecified, model. Structural VARs, however, are not free of problems. For example, the identification restrictions researchers employ are often conventional, are not derived from any model that could potentially be used to interpret the results (see Canova and Pina

(2005)) or may be so weak that they can not separate fundamentally different DGPs (see Faust and Leeper (1997), Cooley and Dwyer (1998)). Moreover, the small scale specifications typically used in the literature are likely to face omitted variable problems. Finally, there are models which can not be recovered when the Wold representation is used to setup a VAR (see Sargent and Hansen (1991) or Lippi and Reichlin (1993)), that may not admit a finite order VAR representation (see Fernandez Villaverde et. al. (2007)) or that, in small samples, are poorly represented with VARs (see Chari et. al (2006)).

Parameter estimation is seldomly the final goal of an applied investigation and conditional forecasting exercises or welfare calculations are generally of interest. For these experiments to be meaningful, one must assess the quality of a model's approximation to the data, both in an absolute sense and relative to a benchmark model. and techniques which are simple, reproducible, effective in comparing the economic discrepancy between the model and the data and informative on the reasons why differences emerge are needed for this purpose. Unfortunately, no existing statistical technique meets all these criteria for two reasons. Traditional econometric methods are unsuited to measure the magnitude of the discrepancy when the model is known to be a false description of the data; statistical criteria give little information on the economic relevance of the discrepancy. Del Negro and Schorfheide (2004) and (2006) have suggested an interesting way to evaluate misspecified models. However, their approach is computationally intensive and not yet tested in coherent experimental designs.

This paper presents a simple approach to validate business cycle models. It employs the flexibility of SVAR techniques against model misspecification and the insight of computational experiments (see e.g. Kydland and Prescott (1996)) to design probabilistic measures of fit which can discriminate among local alternative DGPs and are informative about the economic relevance of the discrepancies. We take seriously the objection that models are at best approximations to portions of the DGP. We are sympathetic to the claim that too little sensitivity analysis is typically performed on calibrated/estimated models and that the reported outcomes may depend on somewhat arbitrary choices. We also pay attention to the fact that identification restrictions typically used in SVAR are often unrelated to the class of models that researchers use to interpret the results.

Our starting point is a class of models which has an approximate state space representation once (log-)linearized around their steady states. We examine the dynamics of the endogenous variables in response to shocks for alternative members of the class using a variety of parameterizations. While magnitude restrictions are often fragile, sign and, at times, shape restrictions are much more robust to the uncertainty we consider. A subset of these robust restrictions is used to identify structural disturbances in the data. Therefore, the minimal set of "uncontroversial" constraints we use to obtain a structural VAR is a collection of robust model-based sign restrictions. We then use the dynamics of unrestricted variables to construct qualitative and quantitative measures of economic discrepancy between a member of the class and the data or between two members of the class. The approach is constructive: if the discrepancy is deemed large at any stage of the evaluation, one can respecify the model and repeat the analysis.

Our methodology is advantageous in several respects. First, it does not require that the true DGP is a member of the class of models we consider. Instead, we only require that a subset of the robust sign restrictions that the selected class implies has a counterpart in the data. Second, our approach does not need the probabilistic structure of the model to be fully specified to be operative. In fact, our procedure can be employed regardless of characteristics or the dynamics induced by other potentially interesting but unmodelled disturbances. Third, by focusing SVAR identification on robust model-based restrictions, our methodology catches several birds with one stone: it de-emphasizes the quest for a good calibration, a difficult task when data is short, unreliable or scarce; it gives content to identification restrictions used in SVARs; it shields researchers against omitted variable biases and representation problems. Fourth, the approach is flexible, it can be used in a limited information or full information mode, and has a few degrees of freedom that can used to make shock identification more or less constrained. Finally, the procedure requires negligible computing power (basically a log-linear solver and a SVAR routine), it is easily reproducible and potentially applicable to a number of interesting economic issues.

We show that our approach can recognize the qualitative features of true DGP with high probability and can tell apart models which are local to each other. It can also provide a good handle on the quantitatively features of the DGP if two conditions are satisfied: identification restrictions are abundant - both in the sense that a large number of variables is restricted and that a large number of restrictions is imposed on a fixed number of variables; the variance signal of the shock(s) one wishes to identify is strong. When this is the case, our approach is successful even when the VAR is misspecified relative to the time series model implied by the aggregate decision rules and the sample is short.

We apply the methodology to study the impact effect of technology shocks on hours worked and the response of consumption to government expenditure shocks, two questions which have received a lot of attention in the recent literature (see e.g. Gali and Rabanal (2004) and Perotti (2007)).

The rest of the paper is organized as follows. The next section describes the methodology. Section 3 examines the ability of the methodology to recognize the true DGP and to distinguish between locally alternative DGPs, both in population and in small samples, in a few controlled experiments. Section 4 presents the two applications. Section 5 concludes.

2 A sign restriction approach to evaluation

It is our presumption that DSGE models, while useful to qualitative characterize the dynamics induced by shocks, are still too stylized to be taken seriously, even as an approximation to part of the DGP of the actual data. Since this misspecification will not necessarily vanish by completing the probabilistic space of the model, we do not try to find parameters that make the model and the data "close" in a quantitative sense.

To describe the details of our approach we need some notation and a few definitions. Let $F(w_t^s(\theta), \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M}) \equiv F^s(\theta)$ be a set of functions, which can be simulated conditional on the structural disturbances ϵ_t, using models in the class $\mathcal{M}$. $F^s(\theta)$ could include impulse responses, conditional cross correlations, etc., and depends on simulated time series $w_t^s(\theta)$, where θ are the structural parameters, and on the parameters of the VAR representation of simulated data, where $\alpha_0(\theta)$ is matrix of contemporaneous coefficients and $\alpha_1(\theta)$ the companion matrix of lagged coefficients. Let $F(w_t, \alpha_0, \alpha_1|u_t) \equiv F(\alpha_0)$ be the corresponding set of functions in the data, conditional on the reduced form shocks u_t.

We take the class of models $\mathcal{M}$ to be broad enough to include submodels with interesting economic features. For example, $\mathcal{M}$ could be one of the standard New Keynesian models used in the macroeconomic literature and the submodels of interest versions where one or more frictions (say, wage stickiness or price indexation) are shut off. The class $\mathcal{M}$ is misspecified in the sense that even if there exists a θ_0 such that $\alpha_0 = \alpha_0(\theta_0)$ or $\alpha_1 = \alpha_1(\theta_0)$ or both, $w_t^s(\theta_0) \neq w_t$ and/or $F(w_t^s(\theta), \alpha_0(\theta_0), \alpha_1(\theta_0)|\epsilon_t, \mathcal{M}) \neq F(w_t, \alpha_0, \alpha_1|u_t)$.

Among all possible functions $F^s(\theta)$, we restrict attention to those $\tilde{F}^s(\theta)$ which are robust: $\tilde{F}_1^s(\theta) \subset \tilde{F}^s(\theta)$ is used for estimation and $\tilde{F}_2^s(\theta) \subset \tilde{F}^s(\theta)$ for evaluation purposes. We assume that $\tilde{F}_1^s(\theta)$ is of dimension $J_1 \times 1$ and that $\tilde{F}_2^s(\theta)$ is of dimension $J_2 \times 1$. $\tilde{F}_1^s(\theta)$ is termed robust if $sgn(F_1^s(\theta_1)) = sgn(F_1^s(\theta_2))$, $\forall \theta_1, \theta_2 \in [\theta_l, \theta_u]$, while $\tilde{F}_2^s(\theta)$ is termed robust if $sgn(F_2^s(\theta_1)|\mathcal{M}_j) = sgn(F_2^s(\theta_2)|\mathcal{M}_j)$, $\forall \theta_1, \theta_2 \in [\theta_l, \theta_u]$, where sgn is the sign of F_i^s; θ_l, θ_u represent the upper and lower range of economically reasonable parameter values and $\mathcal{M}_j \in \mathcal{M}$. Hence, $\tilde{F}_1^s(\theta)$ contains functions whose sign is independent of the submodel and the parameterization, while $\tilde{F}_2^s(\theta)$ contains functions whose sign is independent of a parameterization, given a submodel. The choice of $F_1^s(\theta)$ and $F_2^s(\theta)$ is dictated by the nature of the economic question.

2.1 The algorithm

To keep the presentation simple we describe our approach in the form an algorithm. The procedure involves six specific steps:

1. Find robust implications of the class of models. That is, find the set of functions $\tilde{F}^s(\theta)$, whose sign is independent of θ, and select $\tilde{F}_1^s(\theta)$ and $\tilde{F}_2^s(\theta)$.

2. Use some robust implications to identify shocks in the data. That is, find the set of α_0 that minimizes $I_{[sgnF_1(w_t,\alpha_0,\alpha_1|u_t)-sgnF_1(w_t^s,\alpha_0(\theta),\alpha_1(\theta)|\epsilon_t,\mathcal{M})\neq 0]}$, where $\theta \in [\theta_l, \theta_u]$ subject to $A_0 A_0' = \Sigma_u$, $\alpha_0 = A_0 H$, $HH' = I$ where $I_{[.]}$ is a counting measure, Σ_u the covariance matrix of reduced form disturbances. If there is no α_0 such that $0 \leq \iota \leq I_{[.]}$, some $\iota \geq 0$, stop evaluation.

3. Evaluate the performance **qualitatively** by computing (a) $S_1(\mathcal{M}_j) = \frac{1}{N} \times I_{[sgnF_2(w_t,\hat{\alpha}_0,\alpha_1|u_t)-sgnF_2^s(w_t^s,\alpha_0(\theta),\alpha_1(\theta)|\epsilon_t,\mathcal{M}_j)]=0}$ and /or (b) $S_2(\mathcal{M}_j) = \frac{1}{N} \times I_{[shpF_2(w_t,\hat{\alpha}_0,\alpha_1|u_t)-shpF_2^s(w_t^s,\alpha_0(\theta),\alpha_1(\theta)|\epsilon_t,\mathcal{M}_j)]=0}$, where *shp* is the dynamic shape of F_2, $\hat{\alpha}_0$ are the N values of α obtained in step [2.], and S_1 and S_2 are conditional on model j.

4. Cross validate **qualitatively** different members of the class if needed, i.e. repeat [3.] for each candidate. If one candidate needs to be selected, choose $\mathcal{M}_h, h = 1, 2, \ldots$ that minimizes $S(\mathcal{M}_h) = \sum_{j=1}^{J_1} w_j^1 S_{1j}(\mathcal{M}_h) + \sum_{j=1}^{J_2} w_j^2 S_{2j}(\mathcal{M}_h)$, where $\sum_j w_j^1 + \sum_j w_j^2 = 1$ are weights chosen by the researcher.

5. If the discrepancy in 3.-4. is not too large, continue the validation process **quantitatively**. For example, compute $Pr(F_2^s(\theta) \leq F_2(\hat{\alpha}_0))\ \forall\theta \in [\theta_l, \theta_u]$ or the degree of overlap between $D(F_2^s(\theta))$ and $D(F_2(\alpha_0))$, where the distributions D are computed randomizing over θ and the α_0 found in [.2].

6. Respecify the model if the performance in either 2. or 3.-4.-5. is unsatisfactory. Otherwise, undertake policy analyses, computational experiments, etc. as needed.

The first step of our procedure is explicitly designed to cope with the inherent arbitrariness of calibration procedures: we seek implications which are representative of the class of models we want to evaluate. For example, if the sign of the conditional covariations of output and the nominal interest rate in response to monetary shocks is unchanged when we vary the risk aversion coefficient within a reasonable range, and this is true for the subset of interesting members of $\mathcal{M}$ we consider, we call this a robust implication. Robustness is not generic as many dynamic features are sensitive to the parametrization. Moreover, since models are misspecified, magnitude restrictions are unlikely to hold in the data. Hence, the robust implications we consider take the form of sign restrictions, primarily on the impact period. Also, while both unconditional and conditional dynamics can be used, we find statistics based on the latter more informative.

In the second step we make the class of models and the data share qualitative aspects of their conditional dynamics. This step is easily implementable using the numerical approaches of Canova and De Nicolo (2002) or Uhlig (2005). One can "strongly" or "weakly" identify disturbances, by imposing a large or a small number of robust restrictions, across horizons and/or variables. In line with SVAR practice, we will use a minimal set of restrictions in the identification process. Contrary to standard practices, we derive them explicitly from a class of models and employ only qualitative constraints which are robust. Clearly, some robust restrictions may not hold in the data. In that case, one would either repeat step [2.] imposing an alternative set of robust restrictions, or, if all robust implications are exhausted and no shocks with the required properties found, stop the evaluation process and go back to the drawing board.

The third step is similar to the one employed in computational experiments where some moments are used to estimate/calibrate the structural parameters; others are used to check the performance of the model. Here robust sign restrictions are employed to identify structural shocks; the sign and shape of the dynamic response of unrestricted variables is used to check the quality of the model approximation to the data. We differ from standard practices because, at both stages, we only consider qualitative implications. In the evaluation process we select functions which are robust from the point of view of the model and, ideally, void of measurement error. For example, if a "supply" shock is identified by means of the sign of the joint responses of output and inflation, we could examine whether the sign and the shape of the response of investment or hours to this shock are qualitatively similar in the model and in the data, if the model has robust predictions about the dynamics of these two variables to supply shocks and if their responses can be accurately measured in the data.

At times a researcher may be concerned with the relative likelihood of models which differ in terms of frictions or basic microfundations. If none of the candidates models is discarded after the first three steps of the evaluation procedure, it is possible to qualitatively compare them using qualitative devices such as the sign and shape of selected responses to shocks. A weighted average of counting measures can be used to select the submodel with the smaller discrepancy with the data. If robustness is a concern, pseudo-bayesian averaging, where a scaled version of $S(\mathcal{M}_h)$ is employed as weight, can be employed. Note that candidates could be nested and or non-nested: our method works for both setups.

When the scope of the analysis is to give quantitative answers to certain questions, to undertake conditional forecasting exercises or perform welfare calculations, the quality of the model can be further assessed using probabilistic Monte Carlo methods, i.e. constructing probabilities of interesting events or measures of distance between distributions of outcomes (as e.g. Canova (1995)). The computational cost of this step is minimal since model distributions are obtained in step [1.], and distributions of data outputs in step [3.]. Quantitative evaluation is not a substitute for a qualitative one: candidates can be eliminated and the burden of evaluation reduced if a qualitative check is performed first.

2.2 Discussion

We believe that the procedure is informative about the properties of models and the discrepancy measures provide useful indications on how to reduce the mismatch with the data. For example, shape differences may suggest what type of propagation may be missing while sign differences the frictions/shocks that need to be introduced. Also, contrary to many procedures, the approach permits both sequential and joint identification of sources of shocks.

The approach we propose compares favorably to direct structural estimation and testing of business cycle models for at least two reasons. Classical estimation and inference are asymptotically justified under the assumption that the model used is the DGP of the data. As we have mentioned, such an assumption is probably still too heroic to be entertained, even after frictions, delays restrictions and measurement errors are added to standard constructions. Furthermore, as Canova and Sala (2006) have shown, the mapping between structural parameters and aggregate decision rules in existing models is highly nonlinear and this creates severe identification problems even in large samples.

Both issues are relatively unimportant in our setup. First, the use of robust identification restrictions shields, to a large extent, researchers from the issue of model and parameter misspecification. Furthermore, since we consider only restrictions which are robust to parameter/specification variations, we do not have to take a stand on the relationship between the class of models we consider and the DGP of the data. Second, since our approach does not explicitly use the mapping between structural parameters and aggregate decision rules lack of parameters identification is less of a problem for our approach. Moreover, since the set of α_0's in step [2.] is not necessarily a singleton, the procedure recognizes that with finite samples it may be difficult to uniquely pin down a value of α_0.

SVAR analyses are often criticized because identification restrictions lack a link with the theory that it is used to interpret the results. Since we employ theory based robust sign restrictions, such a problem is absent in our framework. A number of authors have also indicated that another form of subtle misspecification may be present in SVARs. While the literature has cast this problem into an "invertibility" issue (see Fernandez-Villaverde et. al. (2007), Christiano, et. al (2005), Chari et. al (2006) and Ravenna (2007)), to show that our approach is less likely to face it it is more useful to think of it as an omitted variable issue. It is well known that the aggregate decision rules of a log-linearized of a general equilibrium dynamic model have the following state space format

$$\begin{aligned} x_{1t} &= A(\theta)x_{1t-1} + B(\theta)e_t \\ x_{2t} &= C(\theta)x_{1t-1} + D(\theta)e_t \end{aligned} \tag{1}$$

where $e_t \sim iid(0, \Sigma_e)$, x_{1t} are the states, x_{2t} the controls, e_t the exogenous shocks and $A(\theta), B(\theta), C(\theta), D(\theta)$ continuous differentiable function of the structural parameters θ. (1) implies that log-linearized decision rules are members of a larger class of VAR(1) models of the form:

$$\begin{bmatrix} I - F_{11}\ell & F_{12}\ell \\ F_{21}\ell & I - F_{22}\ell \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} e_t$$

Suppose y_{1t} is a vector of variables excluded and y_{2t} a vector of variables included in the VAR and that these vectors do not necessarily coincide with those of the state x_{1t} and control variables x_{2t}. Then, the representation for y_{2t} is

$$(I - F_{22}\ell - F_{21}F_{12}(1 - F_{11}\ell)^{-1}\ell^2)y_{2t} = [G_2 - (F_{21}(1 - F_{11}\ell)^{-1}G_1\ell]e_t \tag{2}$$

Hence, while the model for y_{2t} is an ARMA(∞, ∞), the impact effect of the shocks in the full and the marginalized systems are identical, both in terms of magnitude and sign. Therefore, as long as robust sign restrictions are imposed on impact, this form of misspecification will not affect shock identification. In general, one should derive robust implications integrating out the variables which will be excluded from the VAR (see e.g. Canova et. al. (2006)). In this case, our approach applies with no alterations to this reduced system of equations.

2.3 Comparing our approach to the literature

The methodology we proposed is related to early work by Canova, Finn and Pagan (1994), who tested a RBC model by verifying the unit root restrictions it imposes on a VAR; and to the recent strand of literature who identify VAR shocks using sign restrictions (see Canova and De Nicolo' (2002) or Uhlig (2005)).

Our work is also related to Del Negro and Schorfheide (2004) and (2006), who use the data generated by a DSGE model as a prior for reduced form VARs. Two main differences set our approach apart: these authors condition on one model in their analysis while we consider a general class of models; they examine quantitative restrictions, while we only work with qualitative ones. This focus allows generic forms of model misspecification to be present and vastly extends the range of structures for which validation becomes possible.

Corradi and Swanson (2007) have also suggested a procedure to test misspecified models. Their approach is considerably more complicated than ours, requires knowledge of the DGP and is not necessarily informative about the reasons for the discrepancy between the model and the data. Finally Chari, et. al. (2007) evaluate business cycle models using reduced form "wedges". Relative to their work, we use a structural conditional approach and probabilistic measures for model comparison exercises.

3 The procedure in a controlled experiment

We choose in our exercise a class of New-Keynesian models similar to the one employed by Erceg et. al. (2000) and Rabanal and Rubio Ramirez (2005), which allows for habit in consumption, and for price and wage indexation mechanisms. We choose this class for two reasons. First, several simpler models are nested into the general setup. Second, the structure is flexible, tractable and informative about the properties of our approach. In the first part of the section we ask what are the size and power properties of our procedure in population, when data is generated by different members of this class. The second part describes whether and how these conclusions are altered by sampling uncertainty.

3.1 The class of models

The equilibrium conditions of the prototype economy, where all variables are expressed in log deviations from the steady state, are

$$
\begin{aligned}
\lambda_t &= E_t\lambda_{t+1} + (r_t - \pi_{t+1}) &(3)\\
\lambda_t &= e_t^b - \frac{\sigma_c}{1-h}\left(y_t - hy_{t-1}\right) &(4)\\
y_t &= e_t^z + (1-\alpha)n_t &(5)\\
mc_t &= w_t + n_t - y_t &(6)\\
mrs_t &= -\lambda_t + \gamma n_t &(7)\\
w_t &= w_{t-1} + \pi_t^w - \pi_t &(8)\\
\pi_t^w - \mu_w\pi_{t-1} &= \kappa_w\left[mrs_t - w_t\right] + \beta(E_t\pi_{t+1}^w - \mu_w\pi_t) &(9)\\
\pi_t - \mu_p\pi_{t-1} &= \kappa_p\left[mc_t + e_t^\mu\right] + \beta(E_t\pi_{t+1} - \mu_p\pi_t), \quad e_t^\mu \sim N(0,\sigma_\mu^2) &(10)\\
r_t &= \rho_r r_{t-1} + (1-\rho_r)\left[\gamma_\pi\pi_t + \gamma_y y_t\right] + e_t^r, \quad e_t^r \sim N(0,\sigma_r^2) &(11)\\
e_t^z &= \rho_z e_{t-1}^z + u_t, \quad u_t \sim N(0,\sigma_z^2) &(12)\\
e_t^b &= \rho_b e_{t-1}^b + v_t, \quad v_t \sim N(0,\sigma_b^2) &(13)
\end{aligned}
$$

Equation (3) is the consumption Euler equation: λ_t is the marginal utility of consumption, r_t the nominal interest rate, π_t the price inflation, and e_t^b a preference shock. Equation (4) defines the marginal utility of consumption with external habit formation. The production function is in (5); e_t^z is an exogenous productivity process and n_t hours worked. Real marginal costs mc_t are defined in (6), where w_t is the real wage. Equation (7) gives an expression for the marginal rate of substitution mrs_t. Equation (8) is an identity linking the real wage growth to the difference between nominal wage and price inflation. The wage and price Philips curves arising from Calvo nominal rigidities are in (10) and (9). μ_p and μ_w parameterize the degree of backward-lookingness in price setting and wage setting, respectively; e_t^μ is a price markup shock, and π_t^w wage inflation. The slope of the price Phillips curve is $\kappa_p \equiv \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p}\frac{1-\alpha}{(1-\alpha+\alpha\epsilon)}$ and the slope of the wage Phillips curve is $\kappa_w \equiv \frac{(1-\zeta_w)(1-\beta\zeta_w)}{\zeta_w(1+\varphi\gamma)}$. The central bank adjusts the nominal interest rate r_t according to the rule in (11). The four exogenous processes are driven by mutually uncorrelated, mean zero innovations. The total factor productivity shock e_t^z and the preference shock e_t^b have autocorrelation coefficients ρ_z and ρ_b, respectively. The monetary shock e_t^r and the markup shock $e_t^{\mu p}$ are iid.

It is straightforward to check that at least five different sub-models are nested into this general structure, which we label M1 - a flexible price, sticky wage model ($\zeta_p = 0$), which we label M2; a sticky price, flexible wage model ($\zeta_w = 0$), which we label M3; a flexible price and flexible wage model ($\zeta_p = 0$, $\zeta_w = 0$), which we label M4; a model with no habits ($h = 0$), which we label M5, a model with no indexation ($\mu_p = 0, \mu_w = 0$), which we label M6.

Next, we need to find robust sign restrictions that hold across a wide range of parameter values and for sub-models in the class represented by (3)-(13). We specify a uniform distribution for the unrestricted parameters over an interval, which we choose to be large enough to include theoretically reasonable values, values obtained with structural estimation procedures or used in calibration exercises - see Table 1.

The discount factor β and the markup parameters ϵ and φ are fixed as they are not separately identified - they enter the two Phillips curves as composites, together with the price and wage stickiness parameter, respectively. The range for other parameters are quite standard. For example, the interval for risk aversion coefficient contains the typical values used in the calibration literature (typically 1 or 2), but it also allows higher values which are sometimes used in the asset pricing literature (see e.g. Bansal and Yaron (2004)). Also, we are quite agnostic about the possible values that the habit and the Calvo parameters can take: the range include, roughly, the universe of possible values considered in the literature.

	Parameter	Support
β	discount factor	0.99
ϵ	elasticity in goods bundler	6
φ	elasticity in labor bundler	6
σ_c	risk aversion coefficient	[1.00, 5.00]
γ	inverse Frish elasticity of labor supply	[0.00, 5.00]
h	habit parameter	[0.00, 0.95]
ζ_p	probability of keeping prices fixed	[0.00, 0.90]
ζ_w	probability of keeping wages fixed	[0.00, 0.90]
μ_p	indexation in price setting	[0.00, 0.80]
μ_w	indexation in wage setting	[0.00, 0.80]
α	1 - labor share in production function	[0.30, 0.40]
ρ_r	inertia in Taylor rule	[0.25, 0.95]
γ_y	response to output in Taylor rule	[0.00, 0.50]
γ_π	response to inflation in Taylor rule	[1.05, 2.50]
ρ_z	persistence of productivity	[0.50, 0.99]
ρ_b	persistence in taste process	[0.00, 0.99]

Table 1: Range of parameter distribution.

Given these intervals, we draw a large number of replications, compute impulse responses and examine the sign of the 95 percent response bands at certain horizons. Table 2 reports the outcomes of our analysis when we focus attention on the impact period - the appendix, which plots pointwise response bands obtained at different horizons for the model M1 shows that, for serially correlated shocks, several restrictions hold for a number of horizons. For each shock, table 2 reports six columns, one for each of the models: a '+' indicates robustly positive responses; a '-' robustly negative responses; a '?' responses which are not robust; and 'na' responses which are zero by construction. The variables we report are the real wage (w_t), the nominal rate (r_t), the inflation rate (π_t) and the output gap (y_t), hours worked (l_t).

	Markup shock						Monetary shock						Taste shock						Technology shock					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
r_t	+	+	+	+	+	+	+	+	+	na	+	+	+	+	?	?	+	+	-	-	-	-	-	-
w_t	-	-	-	-	-	-	?	+	-	na	?	?	?	-	?	-	?	?	?	+	?	+	?	?
π_t	+	+	+	+	+	+	-	-	-	na	-	-	+	+	?	?	+	+	-	-	-	-	-	-
y_t	-	-	-	-	-	-	-	-	-	na	-	-	+	+	+	+	+	+	+	+	+	+	+	+
l_t	-	-	-	-	-	-	-	-	-	na	-	-	+	+	+	+	+	+	-	-	-	-	-	-

Table 2: Sign of the impact responses to shocks, different models.

Many of the contemporaneous responses to shocks have robust signs, both across parameterizations and sub-models. For example, positive markup shocks increase the nominal interest rate and inflation, while they decrease the real wage, employment and output on impact with high probability. This pattern is simple to explain: positive markup shocks increase production costs. Therefore, for a given demand, production and employment contract while inflation and the nominal rate increase. In general, taste shocks are the disturbances delivering contemporaneous responses which are less robust across sub-models and the real wage is the variable whose impact response is less robust to shocks in various sub-models. This is because the sign of real wage responses crucially depends on the degree of wage stickiness relative to price stickiness. Since we allow for a wide range of values for these parameters real wages may fall or rise.

The response of the real wage is of particular interest when distinguishing between different submodels in the class. For instance, in model M3 (sticky prices, flexible wages) real wages fall in response to a contractionary monetary shock. With flexible wages, workers are on their labor supply schedule and on impact $w_t = \left(\sigma_c + \frac{\gamma}{1-\alpha}\right) y_t$, so that real wages are procyclical when a monetary shock hits the economy. In model M2 (flexible prices and sticky wages), workers are off their labor supply schedule and from the firm's labor demand schedule, $w_t = -\frac{\alpha}{1-\alpha} y_t$. Hence, real wages are countercyclical in response to monetary shocks with sticky wages and flexible prices. Clearly, one cannot distinguish between sticky price and sticky wage models by measuring the unconditional cyclicality of wages. In each model, there are shocks that make real wages countercyclical and others that make real wages procyclical. However, with sign restrictions on the conditional responses to the real wage, we can shed light on the validity of different submodels.

3.2 Can we recover the true model?

We conduct a few experiments designed to check whether our procedure can correctly recover the sign of certain responses when we endow the researcher with the correct model and a subset of the restrictions shown in table 2.

In the first experiment, we take M2, the flexible price, sticky wage model as our DGP and consider a VAR with the five variables of interest. To avoid singularity, we add one measurement errors to the solution attaching it to the real wage. The parameter of the DGP are in the first column of table 3. We assume that both the model dynamics and the covariance matrix of the reduced form errors Σ are known. We impose various subset of the restrictions and examine whether the impact response of the real wage can be signed with high probability. The behavior of the real wage in this economy is of interest to understand the properties of our procedure since, it is robust across parameterizations in each submodel and, for example, has robustly different signs in models M2 and M3.

We draw a large number of normal, zero mean, unitary 5×5 matrices, use a QR decomposition and construct impact responses using $S * Q$, where S is matrix orthogonalizing the covariance matrix of VAR shocks.

Initially, we impose 16 impact restrictions on output, inflation, hours and the nominal rate and identify all four shocks. We find that 32 of the 10^6 draws satisfy the 16 impact restrictions we impose and that the sign of the impact response of the real wage to markup, monetary, taste and technology shocks has the correct sign in 100, 71, 96 and 100 percent of the cases, respectively. To examine the importance of imposing enough constraints in the identification process, we repeat the experiment by eliminating the contemporaneous restrictions on output. That is, we impose only 12 impact constraints to identify the four shocks. In this case, we find that 278 of the 10^6 draws satisfy the restrictions. For these draws the impact responses of the real wage to markup, monetary, taste and technology shocks have the right sign in 100, 54, 95, 99 per cent of the cases, respectively. Why is there a significant decrease in the percentage of correctly recognized impact signs of the real wage to monetary shocks when we identify all shocks and drop output restrictions? Real wage increases in response to monetary shocks in the model but the magnitude is pretty small. Therefore, unless there are abundant

	Parameter		
β	discount factor	0.99	0.99
ϵ	elasticity in goods bundler	6	6
φ	elasticity in labor bundler	6	6
σ_c	risk aversion coefficient	8.33	8.33
γ	inverse Frish elasticity of labor supply	1.74	1.74
h	habit parameter	0	0
ζ_p	probability of keeping prices fixed	0	0.75
ζ_w	probability of keeping wages fixed	0.62	0
μ_p	backward lookingness price setting	0	0
μ_w	backward lookingness wage setting	0	0
α	1 - labor share in production function	0.36	0.36
ρ_r	inertia in Taylor rule	0.74	0.74
γ_y	response to output in Taylor rule	0.26	0.26
γ_π	response to inflation in Taylor rule	1.08	1.08
ρ_z	persistence of productivity	0.74	0.74
ρ_b	persistence in taste process	0.82	0.82
σ_z	standard deviation of productivity	0.0388	0.0388
σ_μ	standard deviation of markup	0.3167	0.3167
σ_b	standard deviation of preferences	0.1188	0.1188
σ_r	standard deviation of monetary	0.0033	0.0033
σ_{m1}	standard deviation of measurement error 1	0.0001	0.0001

Table 3: Parameter values used in experiments.

restrictions, the estimated impact response of the real wage may marginally fall with certain draws. The fall however is always small.

Next, we examine whether the results are sensitive to the choice of the number of shocks we identify. Intuitively, one should expect to find a larger number of draws satisfying the restrictions when a smaller number of shocks is identified, but the percentage of impact responses of the variable of interest correctly signed will not necessarily increase. Consistent with our a-priori expectations, we find that if, e.g., we identify only technology shocks, about one-fourth of the 10^6 draws satisfy the impact restrictions we impose on output, inflation, the nominal rate and hours but that in only 77 percent of the cases the responses of the real wage are correctly signed. However, no drop in the precision can be found when we identify markup shocks: about that two-third of the 10^6 draws now satisfy the restrictions and in 100 percent of the case the impact response of the real wage is correctly signed.

Why is it that the sign of the real wage responses to markup shocks is always correctly recognized regardless of the number of restrictions we impose? As we will see in details in the next experiment, the variance of the structural shocks matters for the properties of our procedure. From table 3 is evident that markup generate a very strong signal. Therefore, they are easy to identify regardless of the number of restrictions we impose.

The second experiment we run, takes the same model but considers a four variable VAR, which includes output, inflation, nominal rate and the real wage. We fix the parameters of the DGP exactly as in the previous experiment - no measurement error is included here - and still assume that the AR coefficients and the variance covariance matrix of reduced form shocks are known. To identify the four shocks we impose 12 impact restrictions on the response of inflation, the nominal rate and the wage rate and check whether our procedure can correctly sign the impact response of output to each shock. This experiment differs from the previous one in an important aspect: while the previous VAR excluded a state variable - the observed real wage is a contaminated signal of the true one - the current one includes all them. However, given the previous discussion and the fact that we are considering population impact responses, no major changes in the quality of the results are expected.

We run three separate exercises with this specification: (a) we jointly identify all the shocks; we only identify (b) the monetary policy shock or (c) the markup shock. For exercise (b), we allow the variance of the monetary shocks to have different magnitude. Paustian (2007) has shown that what matters is the variance of the shocks one wants to identify relative to the variances of the other shocks matters for identification. The previous experiment seems to indicate that it is the combination of number of restrictions and magnitude of the variance of the shocks that it is crucial. Here we want to examine if this qualification is valid also in the alternative setup we are considering. In particular, we want to know whether real wage response to a monetary policy shock are better identified if variance of the shocks is larger relative to what we have assumed in table 3.

When 12 restrictions on the impact responses are imposed, we find 15 out of 10^6 draws satisfy the restrictions and that the percentage of correctly signed impact output responses is 100, 37, 62, 100 for markup, monetary, taste and technology shocks respectively. When we identify monetary policy shocks only, about one percent of the draws satisfy the three impact restrictions we impose, but the output response is correctly signed in only 35 percent of the cases. This value increases to about 70 (85) percent if the variance of the monetary shocks is multiplied by a factor of 10 (100). Finally, we confirm that markup shocks are much easier to identify than other shocks. In fact, they are obtained in 98 per cent of the draws when only three impact restrictions are imposed and the output response is correctly signed in over 90 percent of the cases, regardless of the variance of the other shocks.

Overall, this set of experiments suggests that our procedure can recognize the qualitative features of the DGP with high probability, when the ideal conditions we consider in this section hold. Nevertheless, three points need to be made. First, when a small number of identification restrictions is used - both in the sense of leaving many variables unrestricted or identifying only one shock - identification becomes weak and, unless the variance of the shock is large, results becomes less favorable. Hence, lacking knowledge about the volatility of structural shocks, it may be dangerous to be too agnostic in the identification process since this may jeopardize the credibility of the conclusions one reaches. Second, the relative strength of the variance signal is crucial for successful inference: the responses of disturbances which are strong and loud are much more easily characterized, regardless of the number of restrictions we impose. Third, and consistent with the theoretical arguments, omitting state variables from the empirical model becomes less crucial when sign restrictions on the impact responses are used for identification and outcomes are evaluated using probabilistic measures.

3.3 Summarizing the features of DGP

So far our analysis has concentrated on the sign of the impact effect of selected variables left unrestricted in the identification process. For many empirical purposes this is focus is sufficient: business cycle theories are typically silent about the magnitude or the persistence of the responses to shocks. At times, however, a more quantitative evaluation is needed. For example, one may be interested in knowing in which percentile of the estimated distribution of responses the true responses lie or whether there exists a location measure that reasonably approximates the true conditional dynamics.

To examine these questions we perform a number of exercises using the flexible price, sticky wage model M2, where we let the real wage be contaminated by iid errors. We identify shocks using a 5 variable VAR using sign restrictions on the impact response of output, inflation, the nominal rate and hours worked and examine the dynamic response of the real wage to the identified shocks for 12 horizons. Again, we assume that both the coefficients of the VAR representation and the covariance matrix of the shocks are known - the only source of randomness is due to identification uncertainty. To be able to measure this uncertainty with some precision, we draw until at least 200 candidates satisfying the restrictions are found - we have checked that with this number of draws identification uncertainty is robustly characterized repeating the experiment a number of time. For these draws, the contemporaneous response of the real wage to markup, technology and taste shocks is correctly signed in 100, 95 and 99 percent of the cases, while the contemporaneous response of the real wage to monetary shocks has the correct sign in 57 percent of the cases.

Figure 1 plots the median and the 95 percent bands of the responses (computed ordering the candidate responses, horizon by horizon, and taking the 2.5, 50 and 97.5 percentile of the distribution) and the true responses. The median is a reasonable although imperfect estimator of the true real wage dynamics in response to shocks. The imperfection comes from the fact that true responses are at times in the tail of the distribution of responses at almost all horizons and at times near the middle of the band (compare wage responses to markup and to technology shocks). Because of this heterogeneity a single location measure must display some bias. Other location measures, such as the arithmetic or the trimmed mean, have similar properties since the distribution of responses at all horizons is roughly symmetric and free of outliers in this experiment.

As a further check of the performance of the median (or the mean) response as estimator for the true responses, we have calculated the population contemporaneous correlation between the true disturbances and disturbances obtained by taking the average median (average) value of the identification matrix. This correlation is computed as follows. The VAR residuals u_t and the true structural residuals e_t are related via $u_t = D\Sigma_e^{0.5} e_t$, where Σ_e is the diagonal covariance matrix of structural shocks and the matrix D comes from the state space representation of the decision rules. Our algorithm delivers for any accepted draw j a matrix Q^j such that $Q^j(Q^j)' = I$. Therefore, a candidate vector of structural shocks satisfies $e_t^j = (SQ^j)^{-1}u_t$, where S is the lower triangular Choleski factor of the VAR residual covariance matrix. Since structural shocks e_t have unitary variances, the correlation between the candidate structural shocks e_t^j and the true structural shocks e_t, $corr(e_t^j, e_t) = (Q^j)^{-1}S^{-1}D\Sigma_e^{0.5}$. Hence, the median correlation is $corr(e_t^{med}, e_t) = (Q^{med})^{-1}S^{-1}D\Sigma_e^{0.5}$ and the average correlation is $N * corr(e_t^A, e_t) = \sum_j (Q^j)^{-1}S^{-1}D\Sigma_e^{0.5}$, where N is the number of accepted draws.

The contemporaneous correlation between true and average recovered shocks of the same type is reasonably high (around 0.6 for all four shocks) but, at times, there is some contamination. For example, the recovered markup shocks have an average correlation of -0.28 with the true taste shocks and the recovered taste shocks have an average correlation of -0.25 with the true technology shocks. In all cases and for all the replications we have run, the highest correlation of the

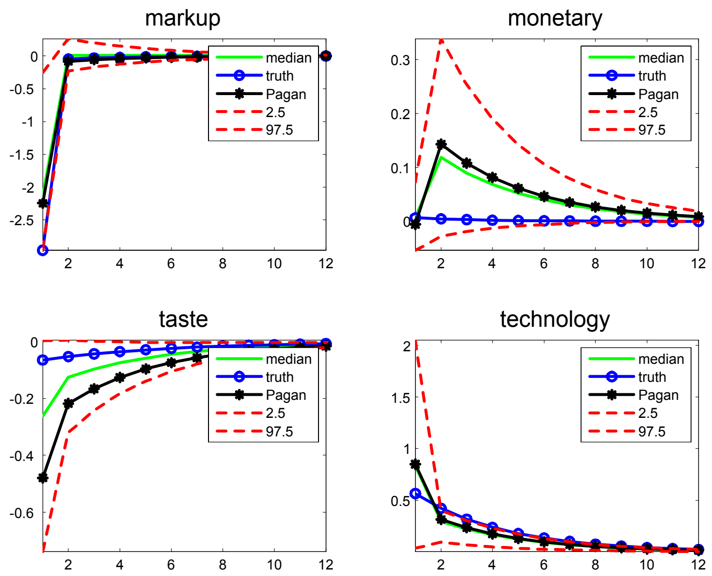


Figure 1: Real wage response to shocks

recovered shocks is always with the corresponding true disturbances. This is good news: since sign restrictions are not designed for this purpose, the fact that they perform well in a quantitative exercise should increase the confidence that researchers have in using them as evaluation devices.

Fry and Pagan (2007) have recently criticized the practice of reporting the median of the distribution of responses as a measure of location when structural disturbances are identified with sign restrictions since the median at each horizon and for each variable may be obtained from different candidate draws and this makes inference difficult. As an alternative, they suggest to use the single identification matrix that comes closest to producing the median impulse response. Figure 1 reports the Fry and Pagan measure, together with true response, the basic median and the 2.5 and 97.5 percentiles. The figure shows that the median response is typically a reasonable measure of the true response. Furthermore, the Pagan median is extremely close to the basic median for markup and technology shocks. These measures differ visibly for monetary and taste shocks, but the Pagan median is not closer to the true response. Thus, while the Pagan median has the attractive property of generating impulse responses that come from a single orthogonal decomposition of the covarianc matrix, it does not necessarily perform better in the sense of getting closer to to the true response. We have also examined the trimmed mean as another summary measure of location. For most controlled experiments, the trimmed mean was very close to the median as the distribution of responses is often symmetric.

We have conducted a number of additional exercises to check whether the performance of location statistics is affected by small changes in the experimental design. We would like to discuss the results obtained when we reduce the variance of markup shocks by 90 percent. As mentioned, markup shocks generate a very strong signal and this makes the identification of other shocks more difficult. By reducing their variance, one should expect the quality of our location measures to improve. Confirming a-priori expectations, we find that it is now easier to recover the other three shocks and for many more candidates the contemporaneous response of the real wage has the correct sign (up about 15 percent). As a consequence, the median becomes a better estimator of the true dynamic responses to monetary, technology and taste shocks.

3.4 Can we exclude alternative models?

The next set of experiments is designed to evaluate whether our procedure is able to eliminate candidate sub-models as potential generators of the data. In the first set of exercises we take M3, the flexible wage, sticky price model, as the DGP and use the parameters in the second column of table 3. We then consider a VAR with 5 variables (real wage, output, inflation, nominal rate and hours) and maintain that the dynamics and the covariance matrix of reduced form shocks are known. Since there are four structural disturbances, we add one measurement error to the real wage to avoid singularity of the covariance matrix. We restrict attention to monetary shocks, which we identify by imposing sign restrictions on the impact response of output, inflation, the nominal rate and hours present in table 2, and ask whether we would be able to exclude that the data were generated by a flexible price, sticky wage version the general model (this is M2), just by looking at the impact response of the real wage which has a different sign in the two specifications.

We draw 10^6 identification matrices and follow the same approach of subsection 3.2. In about 10 percent of the draws the four impact restrictions we impose are satisfied and in over 98 percent of cases, the real wage falls, as the theory predicts. Hence, we are able to exclude that the sticky wage, flexible price version model is the DGP of the data with high probability. To check that this outcome is not due to chance, we have examined two alternative parameterizations of the M3 model, where either the variance of the monetary shocks or of the technology shocks is increased by a factor of ten. When the variance of monetary (technology) shocks is larger about 43 (2) percent of the draws satisfy the restrictions, but in both cases the faction of contemporaneous real wage responses correctly signed exceeds 99 percent.

Why can we recover the sign of the impact response of real wage so well? One possibility is that the responses of the real wage to monetary shocks varies little across parameterizations. Consequently, our success follows from the fact that the real wage is almost always positive in M3, almost always negative in M2 and the area of overlap is small. This does not seem to be the case: the range of real wage impact responses in M3 is relatively large (-0.05, -0.60). Nevertheless, the degree of overlap with the responses produced by M2 is relatively small.

Next, we turn around the null and the alternative hypotheses, that is, we simulate data from a sticky wage, flexible price version of the model and ask whether we can exclude with high probability that the data were generated by the sticky price, flexible wage version of the model. The parameterization we use is in the first column of table 3; the details of the simulation are identical to the previous ones. Once again, the procedure is quite successful: in more than 80 per cent of the draws, the contemporaneous response of the real wage to monetary shocks has the correct sign and at horizon 1 and 2 this percentage increases to more than 94 percent. Furthermore, we find that the average contemporaneous correlation between true and extracted monetary shocks is high (about 95 percent); that the recovered monetary shocks have zero correlation with the true markup, taste and technology shocks; and that the pointwise median response captures well both the magnitude and the shape of the true real wage responses to monetary shocks.

Canova and Sala (2006) and Iskrev (2007) have shown that structural econometric approaches have difficulties in separating sticky price and sticky wage models, because the impulse response based distance function or the likelihood function, are very flat in the parameters controlling price and wage stickiness. Our results suggests that the sign of the impact response of the real wage to monetary shock can recognize very well the nature DGP. Hence, it is comforting to see that our semi-parametric approach can resolve some of identification problems faced by more standard approaches.

Finally, one may want to know whether the ability of our procedure in excluding an alternative sub-model in the same class depends on the parameterization of the DGP. Since the impact response of real wage to monetary shocks is positive in M2 and typically sufficiently large for a wide set of parameters, the parameterization should have little influence on the results. To confirm this, we draw 1000 parameter vectors from the intervals presented in Table 1, except for setting $\theta_w = 0$, and for each draw, we draw 1000 identification matrices. We find that the sign of the real wage on impact is correctly identified on average 99 percent of the times, with a numerical standard error across draws of 4.13. This percentage increases to 99.91 when monetary shocks have larger relative variance (the numerical standard error is 0.8). When the variance of the technology shock is multiplied by a factor of 10, the average percentage of draws satisfying the restrictions is 99.04 (the numerical standard error is 1.27).

To conclude, our procedure has good power in distinguishing models in the ideal situations considered in this subsection: we can exclude potentially relevant candidate DGPs just by using the sign of the impact responses of the real wage, and this is true regardless of the relative size of the variance of the shocks and the exact parameterization of the model. Perhaps more importantly, we can distinguish models in situations where more structural approaches fail.

3.5 How does our approach perform in small samples?

The ideal conditions considered in the previous subsections are useful to understand the properties of the procedure but unlikely to hold in practice. Here we are interested in knowing whether and how conclusions change if the autoregressive parameters and the covariance matrix of the shocks are estimated prior to the identification of the structural disturbances.

To measure sample uncertainty we repeat the experiments we have previously run and consider 200 replications of each experiment. In each replication, we simulate data, keeping the parameters fixed and injected random noise (and measurement error) in the form of normal iid shocks with zero mean and variances reported in table 3. We consider samples with 80, 160 and 500 data points - 20, 40 and 125 years of quarterly data. For each replication we estimate a BVAR, where a close to non-informative conjugate Normal-Wishart prior is used - the results we present are independent of the type of prior we employ. The lag length of the VAR varies: in some cases is the true one, in others it is arbitrarily fixed, in other is estimated using the BIC criteria. We jointly draw from the posterior of the parameters, the covariance matrix of the shocks and the identification matrices until 200 draws satisfying the restrictions are found. We compute pointwise medians and pointwise credible 95 posterior intervals for the variables of interest. For comparison with the true response, obtained from the population VAR representation of the model, we compute the average (or the median) value across replications of the median and the largest interval containing 95 percent of the estimated 95 percent bands at each horizon. We complement these measures with coverage rates - that is, the probability that the true response falls within the estimated credible interval at each horizon - and the probability that responses of certain variables to selected shocks are correctly signed.

We begin from the case where the data is generated from a sticky wage, flexible price version of the general model with one measurement error, and a 5 variable BVAR with output, inflation, the nominal rate, hours, the real wage is used to estimate the dynamics and the covariance matrix of the shocks. The lag length is set to 4 or estimated using a BIC criteria.We identify the four structural shocks imposing sign restrictions on the impact coefficients of output, inflation, the nominal rate and hours to shocks and leave the real wage totally unrestricted. For the sake of presentation, we focus on the dynamics of the real wage when taste and technology shocks hit the economy, as they give the full latitude of estimation results. These responses are in figure 2.

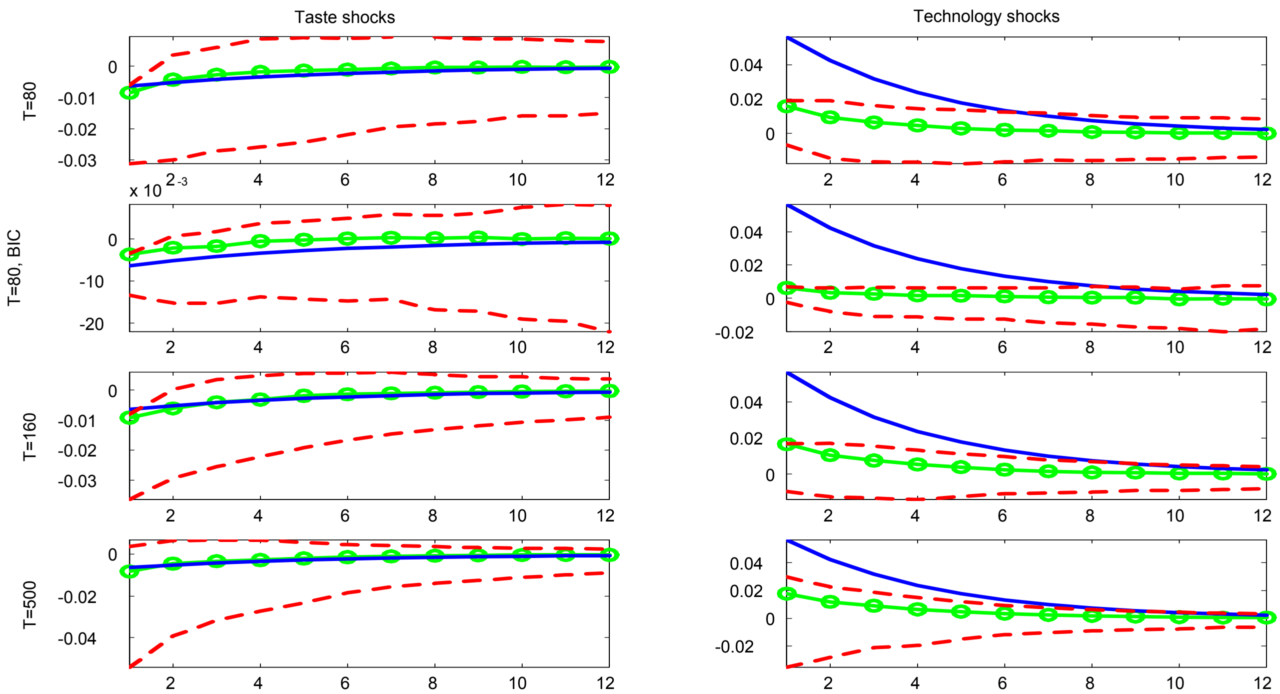


Figure 2: Responses of the real wage to taste and technology shocks

Four features of the figure stand out. First, sample uncertainty is small relative to identification uncertainty: there are some minor differences in the size of the bands obtained as we add observations, but the main features of the estimated dynamics are practically identical. Second, the lag length of the VAR has little consequences on the outcomes of the experiment - this is true even when the sample size is small (see second row of figure 2). Hence, the contribution that longer lags have to the conditional dynamics of the real wage is small and none of the problems highlighted by Fernandez-Villaverde, et. al. (2007) is present here. Third, sample and identification uncertainty compound: the envelope of the bands at each horizon is wide and includes the zero line at every horizon. Taken literally, this means that one can not statistically pin down the dynamics of the real wage to taste and technology shocks. One could make estimation results more precise, for example, by reporting the average of the upper and lower 95 percent credible intervals across replications. However, such a choice has one disadvantage in the presence of outliers or asymmetries: the true responses do not necessarily fall inside the reported band. Fourth, regardless of the sample size, the average median is a good estimator of the shape and of the magnitude of the true wage responses to taste shocks, of the shape of real wage shocks to technology disturbances but not of the magnitude of real wage responses to technology shocks. This asymmetry is due to the fact that the persistence and the unconditional variance of technology shocks are poorly measured.

We have tried to ascertain the size of the sample needed to eliminate the small sample bias in the estimated dynamics of the real wage following technology shocks. We found that only when 500 years of quarterly data are available (200 data points) the true wage response to technology shocks falls into the estimated bands with large probability and, in this case, the median becomes a good estimator also of the magnitude of the true dynamics. Hence, if quantitative comparisons are needed, bias correction techniques such Kilian's (1999), or tighter priors, are needed before the response analysis is performed.

Coverage rates provide little new information. Note that coverage rates for partially identified BVARs will be in general lower than those computed with classical methods and of the nominal rate because of the way identification uncertainty is treated in the two contexts (see Moon and Schorfheide (2007)). We find that in response to taste shocks, the coverage rate is about 70% on impact and increases to about 95% at longer horizons. As the sample size increases, coverage is slightly lower since the estimated bands shrink but the change is small. Coverage rates in response to technology shocks are, as expected, worse in particular at the first few horizons.

In the next experiment we still simulate data from the sticky wage, flexible price version of the model. but we consider a VAR with output, inflation, the nominal rate and the real wage and identify four shocks using impact restrictions on the real wage, inflation and the nominal rate: this leaves the response of output at all horizons completely free. We focus the discussion on the responses of output to markup and monetary shocks, to compare the results with those obtained when only identification uncertainty is present.

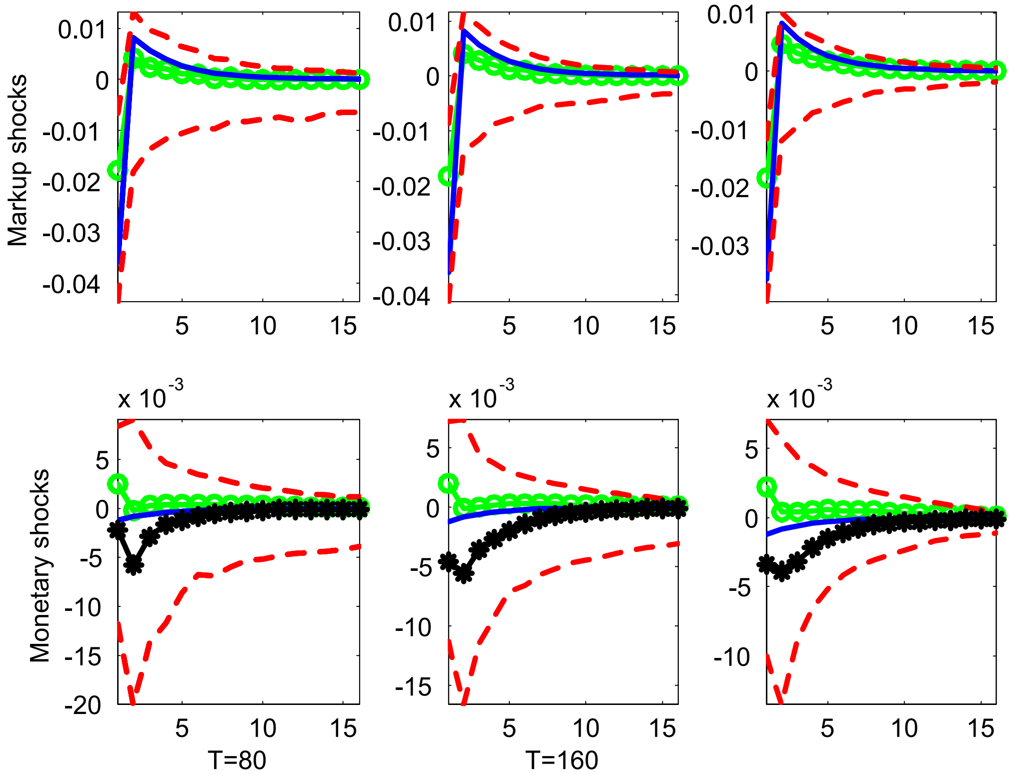


Figure 3: Responses of output to markup and monetary shocks

As figure 3 shows, sample uncertainty adds little to what we already knew: the results obtained with 80, 160 or 500 data points are very similar. Output responses to markup shocks are well estimated, as it was also in the case when sampling uncertainty was absent. Even with 80 data points, all location measures replicate well the shape and the magnitude of these responses. While bands are typically wide and almost always include the zero line, on impact, the envelope of the bands is entirely on one side of zero. The effect of markup shocks on output is well captured, even when a traditional econometric criteria is used.

The magnitude of output responses bands to monetary shocks varies somewhat with the sample size, but in all cases they are large and always contain the zero line. The performance of the average median estimator is invariant to the sample size and, in this instance, not particularly encouraging: while the true output response to monetary shocks is negative on impact, the average median impact (green line) is positive. This is due, in part, to asymmetries in the simulated distribution of the median. For example, the median value of the distribution of the median (starred black line) has the right sign all horizons.

Step	Markup	Monetary	Markup	Monetary	Markup	Monetary
	T=80		T= 160		T=500	
0	92.56	32.03	90.19	32.37	89.85	30.72
1	71.78	50.75	71.31	45.06	68.13	44.01
2	64.09	46.82	69.97	41.53	65.35	40.05
3	66.15	46.54	69.09	38.56	62.20	35.90
4	63.82	45.47	62.78	37.76	58.13	32.09
8	54.60	40.47	54.66	32.21	54.28	27.15
15	53.30	45.04	52.39	37.43	47.90	27.60

Table 4: Probability of correctly signed output responses to shocks.

Table 4, which reports the probability that output responses to shocks are correctly signed at various horizons, confirms what we have learned without sampling uncertainty: the sign of output responses to monetary shocks is poorly measured, regardless of the sample size, while the sign of output responses to markup shocks is well measured even with 80 data points.

Table 4 also confirm that sample uncertainty is, in general, small relative to identification uncertainty. For example, without sampling uncertainty, the impact response of output to monetary shocks was correctly signed in about 37 percent of the cases; this number drops to about 30-32 percent when sampling uncertainty is considered. As before, the number of restrictions and the relative size of the variance of the shocks matter for the performance of the approach but sampling uncertainty has little influence on how these features affect the results.

In standard VARs biases in the estimates of the dynamics of the model are usually of an order of magnitude larger than those in the estimated covariance matrix. This is true also in our case. As shown in tables A.1 and A.2, the true covariance matrix of the shocks is reasonably estimated even with 80 data points but this is not the case for the AR coefficients. For example, the first AR coefficient in the interest rate equation is 1.09 in the model and 0.95 in the data when 80 data points are available, a downward bias of about 15 percent.

In the next experiment we simulate data from the same model we have used in the previous experiment and ask whether with typical samples we could distinguish such a model from one with sticky price, and flexible wages just by looking at the sign of the responses of the real wage to monetary shocks. With only identification uncertainty, we were able to exclude this locally alternative specification with high probability. Does sample uncertainty changes this conclusion?

We estimate a VAR(2) with output, inflation, nominal interest rate and real wage and impose identification restrictions on the sign of the impact response of output, inflation and the nominal rate in response to monetary shocks. This leaves the responses of the real wage completely unrestricted. The table 5 shows that, once again, sample uncertainty adds little: sign restrictions are pretty good tools to distinguish between sticky price and sticky wage models, regardless of the sample size. For example, when T=80, there is only about a 10 percent chance of confusing the two models when we look at the impact response of the real wage - and the joint probability that the response of the real wage is correctly signed at horizons from 0 to 4, is over 60 percent. Table 5 also shows that as the signal produced by the monetary shocks becomes stronger, sample uncertainty matters even less: if monetary shocks were 100 times more volatile than we have assumed, the sign of the impact responses of the real wage would be almost always be correctly recovered. Given these results, it is not surprising to find that the average median real wage responses to monetary shocks is a reasonable estimator of both the magnitude and the shape of the true responses and that coverage rates are everywhere good.

	Basic monetary shocks			10 times larger shocks			1000 times larger shocks		
Horizon	T=80	T=160	T= 500	T=80	T=160	T=500	T=80	T=160	T=500
0	0.88	0.89	0.87	0.87	0.87	0.89	0.98	0.98	0.98
1	0.70	0.78	0.80	0.83	0.86	0.89	0.87	0.92	0.96
2	0.68	0.77	0.78	0.76	0.83	0.87	0.81	0.88	0.94
3	0.62	0.74	0.76	0.68	0.78	0.85	0.71	0.83	0.91
4	0.59	0.71	0.74	0.63	0.73	0.82	0.65	0.77	0.87
8	0.52	0.62	0.66	0.57	0.62	0.69	0.58	0.64	0.70
15	0.54	0.59	0.58	0.52	0.55	0.58	0.55	0.57	0.58

Table 5: Probability of correctly singed wage responses to monetary shocks.

Finally, we simulate data from the flexible wage, sticky price model, and ask whether we could distinguish it from the flexible price, sticky wage model just by looking at the sign of the wage responses to monetary shocks. Relative to the previous experiment, we complicate the setup since real wage is now measured with error, a five variable VAR with output, hours, inflation, the nominal rate and the real wage is used and the lag length of the model is misspecified.

Table 6 presents the probability that wage responses to monetary shocks are correctly signed when the lag length is arbitrarily set to 2, 5 or 10. on average or in the median across replications. Overall, previous conclusions are confirmed and some interesting new aspects emerge. For example, the median of the distribution is superior to the average as an estimator of the true responses, regardless of the sample size and the lag length. Also, increasing the lag length does not necessarily increases the probability that wage responses are correctly signed, particularly at short and medium horizons. Finally, even with 80 observations one can exclude the local alternative model as DGP with almost 80 percent probability.

			Horizon						
Lags	Sample		0	1	2	3	4	8	15
2	80	average	77.22	70.63	68.92	65.55	63.12	51.73	47.67
		median	88.00	75.25	72.25	67.50	66.50	52.00	47.50
	160	average	80.08	77.26	74.36	70.73	67.03	50.81	42.53
		median	93.50	85.00	81.00	76.50	70.50	49.50	42.00
	500	average	82.40	79.78	78.66	76.29	72.91	50.91	39.15
		median	99.25	90.25	88.00	84.00	77.50	51.75	37.25
5	80	average	81.87	73.82	67.42	58.70	53.94	51.2225	49.37
		median	90.00	79.00	71.00	63.50	55.25	52.7500	49.00
	160	average	81.84	76.51	70.91	64.15	60.20	52.3200	47.78
		median	95.25	83.25	74.25	68.25	60.25	52.7500	46.50
	500	average	77.32	76.50	73.61	70.33	63.39	50.0675	44.96
		median	97.00	85.25	79.00	74.00	66.50	51.0000	46.25
10	80	average	79.05	66.04	58.47	51.57	50.95	51.56	52.78
		median	89.50	69.75	63.00	50.25	49.25	52.75	50.90
	150	average	79.21	72.29	65.71	58.87	55.16	50.91	50.90
		median	93.50	78.25	71.50	62.50	56.00	51.00	52.25
	500	average	81.21	78.42	76.72	72.41	66.05	50.99	49.74
		median	99.50	94.00	88.75	78.00	67.50	52.75	50.00

Table 6: Probability of correctly signed wage responses to monetary shocks.

Figure 4, which reports the response to monetary shocks, contains three interesting new features. First, despite being inferior in terms of probability of correct signs, the average median is a very good estimator of the shape and magnitude of the dynamics of real wages in responses to shocks, for all samples and all lag lengths. Second, adding lags to the VAR trades off biases and sampling uncertainty in the estimates. Third, for this experiment estimated bands include the zero line on impact when the sample size is large. Therefore, we need more than 500 data points, before standard statistical inference becomes meaningful.

To summarize, sample uncertainty does not change any of the conclusions we have previously reached: our approach is effective in recovering the qualitative features of the DGP and in excluding local alternative models as potential DGP.

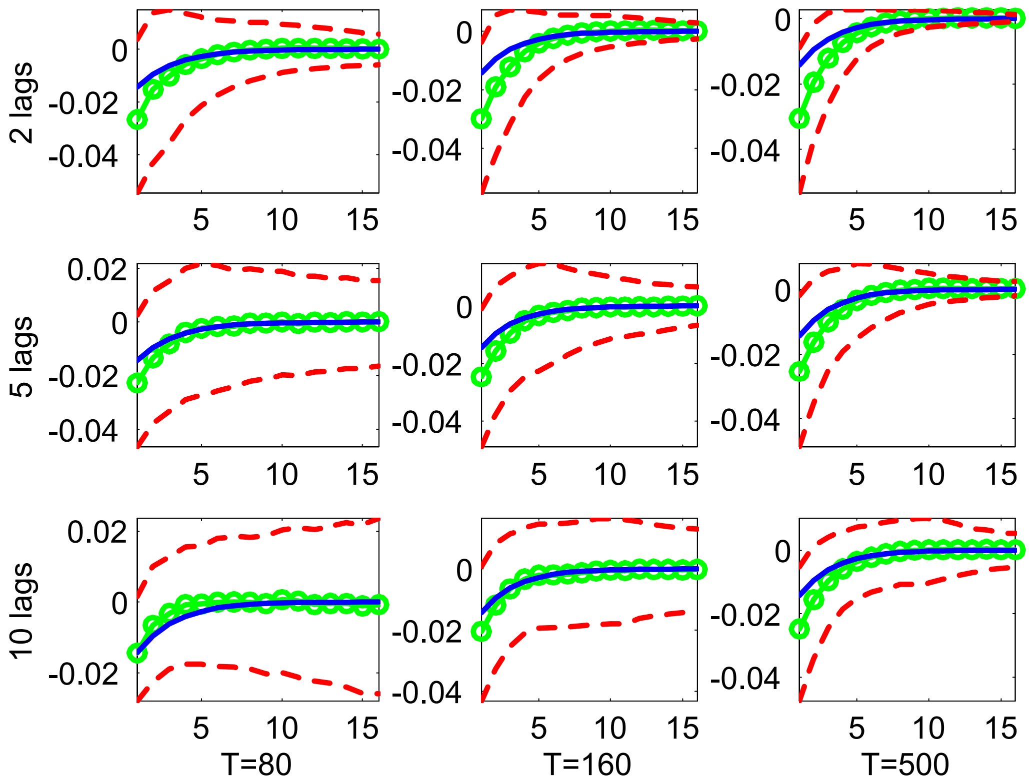


Figure 4: Responses of real wage to monetary shocks

This is true even when the VAR is misspecified relative to the model that has generated the data and in situations where structural estimation approaches fail. In general, a few ingredients are needed to give the methodology its best chance to succeed. First, it is important not to be not agnostic in the identification process: it is probabilistically easier to recognize the DGP if more identification restrictions are used, regardless of the sample size. Second, a stronger variance signal and a sufficiently large number of variables in the VAR help to tell models apart. When these conditions are satisfied the median response, constructed pointwise at each horizon, is a good estimator of the sign and of the shape of the true responses and, in many cases, it can also capture the magnitude of the responses well. Third, if quantitative evaluations are needed, it is important to eliminate biases in estimated VAR coefficients prior to identification of shocks. Absent this correction, the median may become a poor estimator of the dynamics induced by shocks.

Our experiments also show that standard inference is problematic: credible 95 percent intervals tend to be large. Given that sign restrictions produce partially identified models, expecting the same degree of estimation precision as for exactly identified models is foolish. Since the size of the bands is inversely proportional to the number of robust identification restrictions one imposes, and identification uncertainty rather than sample uncertainty dominates, standard statistical analysis may be meaningful only if the identification process is strenghtened by adding as many sign restrictions as possible. Alternatively, one should consider much smaller uncertainty bands, say 68 percent bands or interquartile ranges. The DGP used in this section does not allow much latitude as far as identification restrictions are concerned, unless the intervals for the structural parameter are strongly restricted. Since other DGPs may feature similar problems, considering smaller uncertainty bands should probably be the preferred choice. In general, when identification uncertainty is present, probabilist statements are more informative about the features of the DGP than asymptotically-based standard normal tests.

4 Two examples

4.1 Hours and technology shocks

There has been considerable debate in the literature concerning the sign of the responses of hours to technology shocks. While the debate has often been cast into a RBC vs. New-Keynesian transmission (see Rabanal and Gali (2004) and McGrattan (2004)), researchers have recently start distinguishing various types of technology shocks (Fisher (2006)) and offer alternative (Shumpeterian) explanations of the evidence (see Canova, et. al. (2006)). Rather than entering this controversy, this subsection concentrates on three more narrow questions. First, what kind of hours dynamics are generated by different types of technology shocks? Second, which type of technology shock drives hours fluctuations most? Third, how do technology shocks obtained with long run restrictions relate to those extracted with impact sign restrictions?

To address these questions we use as our prototype class of models, what is considered the benchmark for policy analysis and forecasting in the literature (see Christiano, et. al. (2005) and Smets and Wouters (2003)). This class features sticky nominal wage and price setting, backward wage and inflation indexation, habit formation in consumption, investment adjustment costs, variable capital utilization and fixed costs in production. The log-linearized version of the general model can be characterized as follows. The aggregate demand block is:

$$y_t = c_y c_t + i_y i_t + g_y e_t^g \tag{14}$$

$$c_t = \frac{h}{1+h} c_{t-1} + \frac{1}{1+h} E_t c_{t+1} - \frac{1-h}{(1+h)\sigma_c}(R_t - E_t \pi_{t+1}) + \frac{1-h}{(1+h)\sigma_c}(e_t^b - E_t e_{t+1}^b) \tag{15}$$

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} E_t i_{t+1} + \frac{\phi}{1+\beta} q_t - \frac{\beta E_t e_{t+1}^I - e_t^I}{1+\beta} \tag{16}$$

$$q_t = \beta(1-\delta) E_t q_{t+1} - (R_t - \pi_{t+1}) + \beta r^* E_t r_{t+1} \tag{17}$$

Equation (14) is the aggregate resource constraint. Total output, y_t, is absorbed by consumption, c_t, investment, i_t, and exogenous government spending, e_t^g. Equation (15) is a dynamic IS curve. e_t^b is a preference shock, σ_c is the coefficient of relative risk aversion and h the coefficient of external habit formation. The dynamics of investment are in equation (16). The parameter ϕ represents the elasticity of the costs of adjusting investments, $q_t$is the value of existing capital, e_t^I a shock to the investment's adjustment cost function and β the consumers discount factor. The arbitrage condition for the value of capital is in equation (17): the current value of the capital stock depends positively on its expected future value and its expected return, and negatively on the ex ante real interest rate, r_t. The aggregate supply block is:

$$y_t = \omega(\alpha K_{t-1} + \alpha \psi r_t + (1-\alpha) l_t + e_t^x) \tag{18}$$

$$k_t = (1-\delta) k_{t-1} + \delta i_t \tag{19}$$

$$\pi_t = \frac{\beta}{1+\beta\mu_p} E_t \pi_{t+1} + \frac{\mu_p}{1+\beta\mu_p} \pi_{t-1} + \kappa_p mc_t \tag{20}$$

$$w_t = \frac{\beta}{1+\beta} E_t w_{t+1} + \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta} E_t \pi_{t+1} - \frac{1+\beta\mu_w}{1+\beta} \pi_t + \frac{\mu_w}{1+\beta} \pi_{t-1} - \kappa_w \mu_t^W \tag{21}$$

$$l_t = -w_t + (1+\psi) r_t + k_{t-1} \tag{22}$$

Equation (18) is the aggregate production function. In equilibrium ψr_t equals the capital utilization rate and e_t^x is a neutral shock to total factor productivity. Fixed costs of production are represented by the parameter ω and α is the capital share. The capital accumulation is in (19). Equation (20) links inflation to marginal costs, $mc_t = \alpha r_t + (1-\alpha) w_t e_t^x + e_t^{\mu p}$. The parameter $\kappa_p = \frac{1}{1+\beta\mu_p} \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}$, is the slope of the Phillips curve and depends on ζ_p, the probability that firms face for not being able to change prices in the Calvo setting. The parameter μ_p determines the degree of price indexation and $e_t^{\mu p}$ is a markup shock. Equation (21) links the real wage to expected and past wages and inflation and to the marginal rate of substitution between consumption and leisure, $\mu_t^W = w_t - \sigma_l l_t - \frac{\sigma_c}{1-h(c_t - hc_{t-1})} - e_t^{\mu w}$, where σ_l is the inverse of the elasticity of hours to the real wage and $e_t^{\mu w}$ a labor supply shock and $\kappa_w = \frac{1}{1+\beta}\frac{(1-\beta\zeta_w)(1-\zeta_w)}{\left(1+\frac{(1+\varepsilon^w)\sigma_l}{\varepsilon^w}\right)\zeta_w}$. Equation (22) follows from the equalization of marginal costs. It implies that labor demand depends negatively on the real wage and positively on the rental rate of capital.

Monetary policy is assumed to be conducted according to

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^R \tag{23}$$

where ε_t^R is a monetary policy shock.

Equations (14) to (23) define a system of 10 equations in ten unknowns, $\pi_t, y_t, c_t, i_t, q_t, l_t, w_t, k_t, r_t$, and R_t. Two additional interesting variables can be generated: the ex-post real rate RR_t and the productivity-wage gap ($gap_t = \frac{y_t}{l_t} - w_t$). The model features seven exogenous disturbances: neutral technological, e_t^x, investment-specific, e_t^I, preference, e_t^b, government spending, e_t^g, monetary policy, e_t^R, and price $e_t^{\mu p}$ and labor supply shocks, and $e_t^{\mu w}$. The vector $S_t = [e_t^x, e_t^I, e_t^b, e_t^g, e_t^R, \varepsilon_t^{\mu p} \varepsilon_t^{\mu w}]'$, is parametrized as:

$$\log(S_t) = (I - \boldsymbol{\varrho})\log(\overline{S}) + \boldsymbol{\varrho}\log(S_{t-1}) + V_t \tag{24}$$

where V is a vector of white noises with diagonal covariance matrix Σ_v, $\boldsymbol{\varrho}$ is diagonal with roots less than one in absolute value and $\overline{S}$ is the mean of S.

The model period is one quarter. We split the parameter vector $\theta = (\theta_1, \theta_2)$, where $\theta_1 = (\beta, \pi^{ss}, \Sigma_v)$ are the parameters fixed to a particular value - we fix calibrate them to the posterior mean estimates of Smets and Wouters (2003) - while θ_2 are the parameters which are allowed to vary. Table 7 gives the intervals for θ_2. Note that these ranges are looser than the prior intervals considered in the Bayesian estimation of this model. The range for the investment adjustment cost parameter requires a bit of discussion. Unless this parameter is small, positive investment shocks increase investment too much relative to output, making inflation increase.

We summarize the sign of the responses of the variables of the model to the seven shocks table 8. As in section 3 a '+' indicates a robustly positive sign, a '-' a robustly negative sign and a '?' a sign which is not robust when considering 68 percent simulation bands. Four features of the table allow us to identify the four potential sources of technological improvement (neutral, investment

σ_c	risk aversion coefficient	[1,6]
h	consumption habit	[0.0,0.8]
σ_l	inverse labor supply elasticity	[0.5,4.0]
ω	fixed cost	[1.0,1.80]
$1/\phi$	adjustment cost parameter	[0.0001,0.002]
δ	capital depreciation rate	[0.015,0.03]
α	capital share	[0.15,0.35]
$1/\psi$	capacity utilization elasticity	[0.1,0.6]
g_y	share of government consumption	[0.10,0.25]
ζ_p	degree of price stickiness	[0.4,0.9]
μ_p	price indexation	[0.2,0.8]
ζ_w	degree of wage stickiness	[0.4,0.9]
μ_w	wage indexation	[0.2,0.8]
ε^w	steady state markup in labor market	[0.1,0.7]
γ_R	lagged interest rate coefficient	[0.2,0.95]
γ_π	inflation coefficient on interest rate rule	[1.1,3.0]
ρ_y	output coefficient on interest rate rule	[0.0,1.0]
$\boldsymbol{\varrho}_i$	persistence of shocks $i = 1, \ldots, 7$	[0,0.9]

Table 7: Range of parameter values.

specific, markup and labor supply shocks). First, these shocks increase output and decrease inflation on impact while the other three shocks produce positive comovements of these variables. Second, investment specific shocks make consumption growth fall on impact - the impact response of consumption to the other supply shocks is positive. Third, the impact response of the growth rate of the gap measure is positive in response to technology shocks and negative in response to markup shocks. Fourth, real wage growth falls in response to supply and investment shocks and increases in response to the other two supply shocks.

We use the impact restrictions on output growth, inflation, consumption growth, the productivity-wage gap growth, and the real wage growth to identify the four shocks of interest in the data. These restrictions are satisfied also in submodels of the class, i.e. models with no habit, full utilization, log consumption or linear leisure in utility, no wage stickiness or indexation, no wage and price stickiness, etc. Therefore, they are representative of the class of models we are interested in studying. Note also, that hours robustly fall in response to neutral shocks and robustly increase in response to the other three technology shocks and that these restrictions hold all the submodels we have considered, except when

	Neutral	Monetary	Taste	Investment	Markup	Labor supply	Goverment
Δy_t	+	+	+	+	+	+	+
Δc_t	+	+	+	-	+	+	-
π_t	-	+	+	-	-	-	+
Δgap_t	+	-	-	?	-	+	-
Δw_t	+	+	+	-	+	-	?
Δl_t	-	+	+	+	+	+	+
R_t	?	-	+	+	?	?	+
LP_t	+	-	-	-	?	-	-
i_t	+	+	-	+	+	+	-
u_t	?	+	+	?	+	?	?
RR_t	?	-	?	?	?	?	+

Table 8: Sign of the impact responses to shocks.

investment adjustment costs are set to zero [1].

Figures 5 and 6 report the median and the posterior 68 credible interval at horizons from 0 to 20 for the responses of hours and for the share of hours volatility explained to various technology shocks. It is clear that, while the sign of the responses to labor supply and markup shocks is well estimated, the one to the other two technology shocks is not. Nevertheless, the median of the bands suggests that hours growth instantaneously falls in response to neutral shocks and instantaneously increases in response to the other three shocks.

The four technology shocks explain, in the median, about 50 percent of the forecast error variance of hours growth at horizons varying from 4 to 20 quarters, but the uncertainty is large. Relatively speaking, labor supply and markup shocks explain the largest portion of the hours growth variability at short horizons and neutral shocks the larger portion from horizon six on. Interestingly, investment shocks appear to be a minor contributor of the hours growth volatility at all horizons.

It turns out that technology shocks extracted using long run restrictions and a bivariate VAR with hours and labor productivity are correlated with the neutral and the labor supply shocks we obtain. Interestingly, while the short run correlation with neutral shocks is the largest, the medium run correlation with labor supply shocks is the most significant one.

In sum, the class of models we consider is broadly consistent with the data.

[1] This robustness implies, for example, that the usual RBC vs. New-Keynesian discussion is somewhat sterile. Flexible and sticky price versions of the model imply that hours fall in response to neutral technology shocks, unless investment adjustment costs are set to zero.

Nevertheless, the data speaks loud only for markup and labor supply shocks, both of which seems to matter for hours growth variability only in the short run.

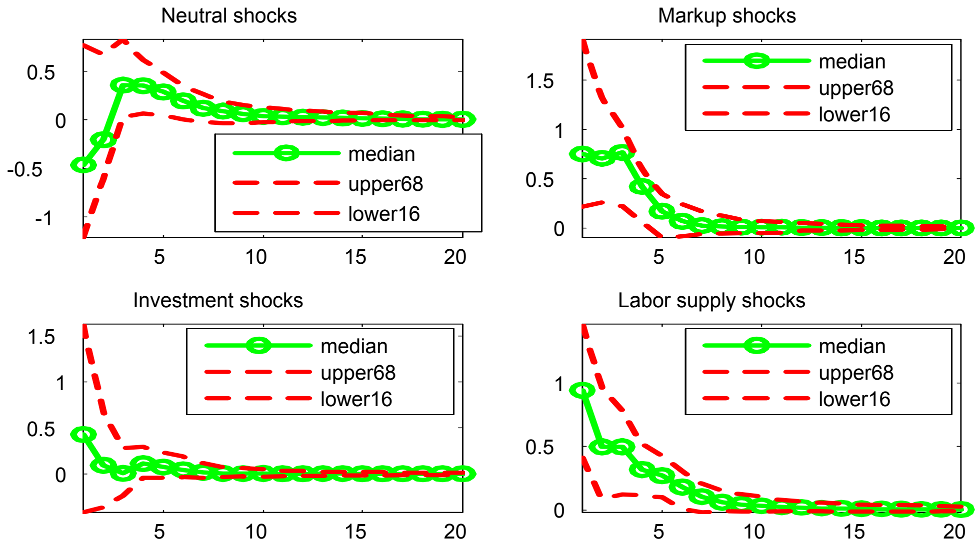


Figure 5: Responses of hours to technology shocks

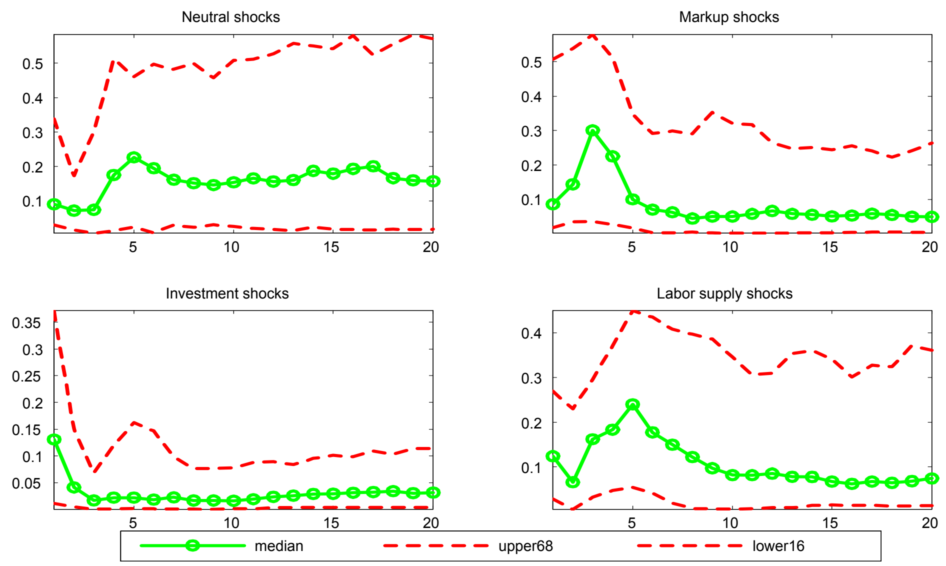


Figure 6: Share of hours volatility explained by technology shocks

4.2 Does consumption increase in response to government expenditure shocks?

The impact response of consumption to government expenditure shocks is controversial. While a portion of the literature suggests that government consumption raises private consumption (see Perotti (2007), crowding out of private consumption is hard to exclude. Gali et al. (2007) suggested that sticky prices and non-ricardian consumers can bring about a simultaneous rise in output and consumption in response to a government spending shock. The log-linearized conditions of the model used by Gali et al. (2007) are the following

$$
\begin{aligned}
q_t &= \beta E_t q_{t+1} + [1-\beta(1-\delta)]\, E_t r_t^k - (r_t - E_t \pi_{t+1}) && (25)\\
i_t - k_{t-1} &= \eta q_t && (26)\\
k_t &= \delta i_t + (1-\delta) k_{t-1} && (27)\\
c_t^o &= c_{t+1}^o - (r_t - E_t \pi_{t+1}) && (28)\\
c_t^r &= \frac{1-\alpha}{\mu \gamma_c}(w_t + n_t^r) - \frac{1}{\gamma_c} t_t^r && (29)\\
w_t &= c_t^j + \phi n_t^j \qquad\qquad j = o, r && (30)\\
r_t^k &= x_t + e_t^z + (1-\alpha)(n_t - k_{t-1}) && (31)\\
w_t &= x_t + e_t^z - \alpha(n_t - k_{t-1}) && (32)\\
y_t &= e_t^z + (1-\alpha) n_t + \alpha k_{t-1} && (33)\\
y_t &= \gamma_c c_t + \gamma_i i_t + g_t && (34)\\
\pi_t - \mu_p \pi_{t-1} &= \kappa(x_t + e_t^u) + \beta(E_t \pi_{t+1} - \mu_p \pi_t) && (35)\\
r_t &= \rho_r r_{t-1} + (1-\rho_r)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^r && (36)\\
b_t &= (1+\rho)[(1-\phi_b) b_{t-1} + (1-\phi_g) e_t^g] && (37)\\
t_t &= \phi_b b_{t-1} + \phi_g e_t^g && (38)
\end{aligned}
$$

Equations (25)-(26) describe the dynamics of Tobin's Q and its relationship with investments i_t. The loglinearized accumulation equation for capital k_t is in equation (27). Equation (28) is the Euler equation for consumption c_t^o of optimizing agents. Consumption of rule of thumb agents c_t^r is determined by their labor income from supplying n_t^r hours of labor at wage w_t net of paying taxes t_t^r as in equation (29). With flexible labor markets, the labor supply schedule for each group is given in equation (30). Cost minimization implies optimality conditions (31) and (32) where x_t is real marginal cost, e_t^z total factor productivity and r_t^k the rental rate of capital. Output is produced according to a constant returns to scale technology as in (33). Market clearing requires that output be absorbed by aggregate consumption c_t, investment i_t and government spending e_t^g.The new Keynesian Phillips curve is in equation (35) where e_t^u is an iid markup shock and μ_p parameterizes the degree of indexation. The central bank conducts monetary policy according to a Taylor rule that relates the nominal interest rate r_t to inflation π_t and output y_t plus a monetary policy shock e_t^r. The government budget constraint together with the fiscal rule gives rise to equation (37) where b_t denotes bonds. The fiscal rule is given by the last equation. The share of rule of thumb agents is denoted by λ. Aggregation implies that $c_t = \lambda c_t^r + (1-\lambda)c_t^o$ and $n_t = \lambda n_t^r + (1-\lambda)n_t^o$.The interesting feature of this model is that if the fraction of rule of thumb consumers is large and prices are sticky enough, a government spending shock can bring about a rise in consumption and output. We use U.S. data and robust sign restrictions derived from this model to test for the presence of rule of thumb consumers. To check for robust model implications, we draw structural parameters from the following intervals.

symbol	description	interval
λ	share of rule of thumb agents	[0.00,0.90]
δ	depreciation of capital	[0.00,0.05]
α	capital share	[0.30,0.40]
θ	price stickiness	[0.00,0.90]
ϕ	inverse of labor supply elasticty	[0.00,5.00]
η	elastictiy of investment	[0.50,2.00]
ρ_r	inertia in monetary policy	[0.00,0.90]
γ_π	policy response to inflation	[1.05,2.50]
γ_y	policy response to output	[0.00,0.50]
μ_p	indexation in price setting	[0.00,0.80]
ϕ_b	fiscal rule response to bonds	[0.25,0.40]
ϕ_g	fiscal rule response to expenditure	[0.05,0.15]
ρ_g	AR(1) parameter gov. spending	[0.50,0.95]
ρ_t	AR(1) parametr productivity	[0.50,0.95]
μ	gross monopolistic markup	[1.10,1.30]
γ_g	steady state spending share in output	[0.15,0.20]

The range for most of the parameters is the same as in the baseline model in section 3. For the fiscal rule parameters we choose an interval centered around the calibrated values in Gali et al. (2007). We draw 10^5 sets of structural parameters from the uniform distribution and keep only those draws for which a determinate rational expectations equilibrium exists. Determinacy is obtained in about 75 percent of the draws. The analysis shows that on impact, the follows signs are robust.

	markup	monetary	spending	technology
r	?	?	+	-
w	-	-	?	?
π	?	-	+	-
y	-	-	+	+
l	-	-	+	-
i	?	?	-	+
c	-	-	?	+

Before take the model by Gali et al. (2007) to U.S. data, we examine how our approach fares with artificial data generated by the model. The model is driven by four shocks: government spending, markup shocks, monetary shocks and technology shocks. We take spending shock and technology shocks to be autocorrelated with AR(1) coefficients set to 0.9. The markup and monetary shocks are taken to be iid. We calibrate the model in the same way that Gali et al. (2007) do in their baseline calibration. For the controlled experiment, we set the standard deviations of monetary shocks to 0.025, of the markup shock to 0.3, of the government spending shock to 0.1 and of total factor productivity to 0.07. We assume the researcher observes data on hours, investment, consumption and inflation. We will assume again that the population VAR representation of these variables is known. We first take as the true DGP the model without any rule of thumb consumers, $\lambda = 0$, and ask if minimal sign restrictions can recover that consumptions falls in response to a government spending shock. We impose that government spending shocks increase hours and inflation and crowd out investment on impact. Since such a pattern could also be induced by negative technology shocks we jointly identify both types of shocks by imposing that a positive technology shock reduces hours and inflation, increases investment, and increases consumption.

Figure 7 shows that our approach captures well the fall in consumption the model generates. On impact 100 percent of the accepted draws have consumption falling. Furthermore, the median identified response of consumption tracks the actual response almost perfectly. Recall that no restriction was imposed on the response of consumption to government spending shocks. Hence, the method works well at pointing towards an absence of rule of thumb consumers.

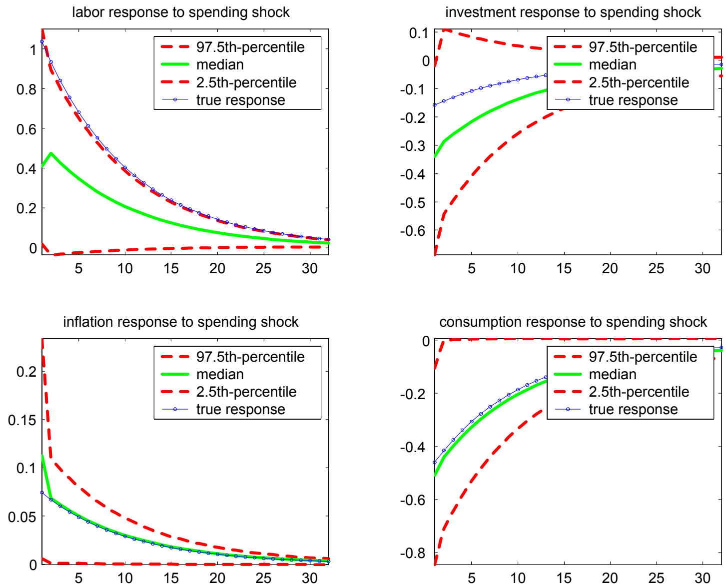


Figure 7: Identified impulse responses to spending shock with $\lambda = 0$.

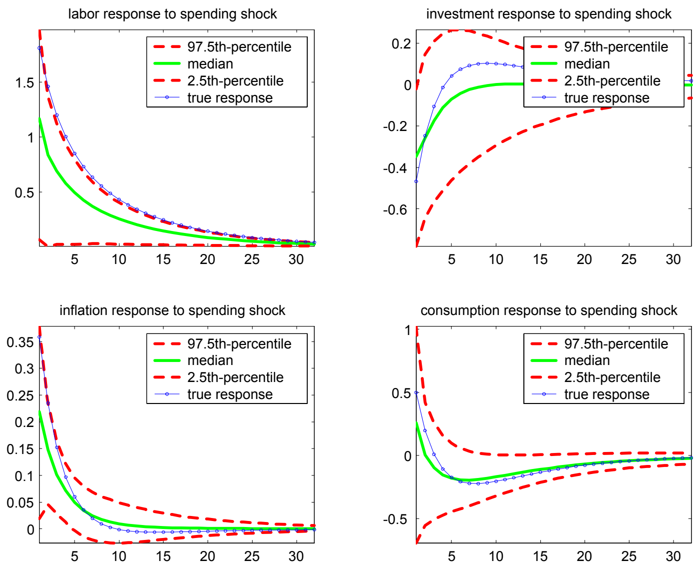
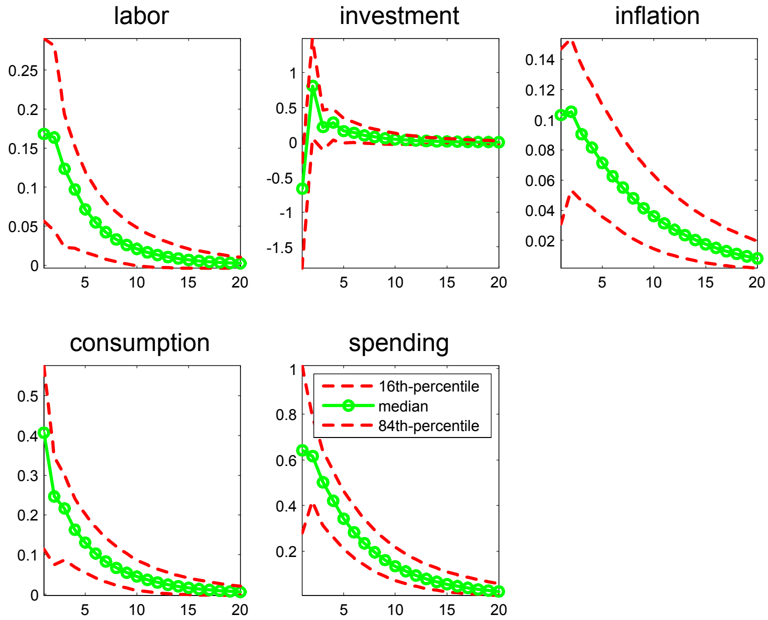


Figure 8: Identified impulse responses to spending shock with $\lambda = 0.8$.

Next, we turn to the case where true DGP features a large enough fraction of rule of thumb consumers, such that consumption clearly rises in response to a spending shock. We set $\lambda = 0.8$ and impose the same restrictions on the population VAR as before. In this case, the identified set includes both positive and negative responses, but consumption rises on impact in about 70 percent of the accepted draws and the median response is again reasonably close to the true response. Given that the procedure performs well in the controlled experiments and we now test for the presence of rule of thumb consumers in U.S. data.



Response to a government spending shock in U.S. data 1954-2007

We estimate a 5 variable BVAR on quarterly U.S. data from 1954-2007 obtained from the FRED database. We estimate the BVAR with an Normal Inverted-Wishart prior. The lag length is chose according to BIC. Our measure of government spending is government consumption expenditures and gross investment (federal, state and local). Consistent with the analysis above, we include hours worked in the nonfarm business sector, consumption, investment, and GDP inflation. All variables enter in logs and first differences, except inflation that is in log levels. We identify government spending shocks by imposing that they raise inflation, governent spending growth, hours growth, and lower investment growth. In addition, we identify technology shocks by imposing that they lower inflation, hours growth, and raise investment growth and consumption growth These restrictions are imposed only on impact. We jointly draw from the posterior and the orthonormal matrices until 2000 draws are found that satisfy the restrictions.

As the above figure shows, the small number of imposed restrictions clearly point to a rise in consumption growth following a government spending shock. The crowding out effect on investment growth appears to be shortlived, whereas the responses of hours growth, consumption growth and inflation are more persistent. Overall, the data is consistent with the key implications of the model by Gali et al. (2007) where consumption rises in response to a government spending shock. This finding is qualitatively unchanged if we do not identify the technology shock along with the government spending shock. When only government spending shocks are identified, the 68 percent credible set contains zero at almost all horizons. However, the median response still indicates a rise in consumption growth. Overall, the data seems to favor models with non-Ricardian consumers.

5 Conclusions

This paper presents a simple methodology based on sign restrictions to examine the validity of business cycle models. The approach employs the flexibility of SVAR techniques against model misspecification and the insight of computational experiments to design probabilistic measures of discrepancy which can discriminate among local alternative DGPs and are informative about the economic relevance of the discrepancies.

Our starting point is a class of models which has an approximate state space representation once (log-)linearized around their steady states. We examine the dynamics of the endogenous variables in response to shocks for alternative members of the selected class using a variety of parameterizations. A subset of these robust restrictions is used to identify structural disturbances in the data. We then use the dynamics of unrestricted variables to construct qualitative and quantitative measures of economic discrepancy between a member of the class and the data and between two members of the class.

Our approach can recognize the qualitative features of true DGP with high probability and it can tell apart models which are local to each other. It can also provide a good handle on the quantitatively features of the DGP if two conditions are satisfied: identification restrictions are abundant; the variance signal of the shock(s) one wishes to identify is strong. In this case, our approach is quantitatively successful even when the VAR is misspecified relative to the time series model implied by the aggregate decision rules and the sample is short.

Our methodology is advantageous in several respects. First, it does not require the true DGP to be a member of the class of models we consider. Second, it does not need the probabilistic structure of the model to be fully specified to be operative. Third, by focusing SVAR identification on robust model-based restrictions, our methodology de-emphasizes the quest for a good calibration and shields researchers against omitted variable biases and representation problems. Fourth, the approach is flexible, it can be used in a limited information or full information mode and require negligible computer time.

We show by means of two examples that the methodology can be very useful to empirical characterize the responses to shocks and help to identify which theories are more relevant to explain the data. Recent work by Dedola and Neri (2007) and Pappa (2005) show that the approach has a considerable number of applications and can be very useful for applied researchers.

References

Bansal, R, and Yaron, A (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles", *Journal of Finance*, LIX, 1481-1509.

Braun, P. and S. Mittnik (1993): "Misspecification in VAR and their effects on impulse responses and variance decompositions, *Journal of Econometrics*, 59, 319-341.

Canova, F. (1995): "Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies", *International Economic Review*, 36, 477-501.

Canova, F. (2004): "Monetary policy and the evolution of the US economy", available at http://www.crei.cat/people/canova/papers.html.

Canova, F. and De Nicoló, G. (2002): " Money Matters for Business Cycle Fluctuations in the G7", *Journal of Monetary Economics*, 49, 1131-1159.

Canova, F. and Pina, J. (2005): "What VARs tell us about DSGE models?", in C. Diebolt, and Krystou, C. (eds.), *New Trends In Macroeconomics*, Springer Verlag, New York, NY.

Canova, F. and Sala, L. (2006): Back to square one: identification issues in DSGE models, ECB working paper.

Canova, F., Finn, M. and Pagan, A., (1994) "Evaluating a RBC model" in C. Hargraves (ed.) *Non Stationary Time Series Analysis and Cointegration*, Oxford: Oxford University Press.

Canova, F, Lopez Salido, D. and Michellcci, C, (2006): Shumpterian Technology Shocks, CEMFI manuscript.

Chari, V.V., Kehoe, P. and McGrattan, E. (2006): A critique of SVAR using Real Business Cycle theory, Fed of Minneapolis, working paper 631.

Chari, V.V., Kehoe, P. and McGrattan, E. (2007): "Business cycle accounting", *Econometrica*.

Christiano, L., Eichenbaum, M. and Evans, C. (2005) Nominal Rigidities and the Dynamic Effects of a shock to Monetary Policy, *Journal of Political Economy*, 113, 1-45.

Christiano, L., Eichenbaum, M. and Vigfusson, R. (2006) Assessing structural VARs, NBER working paper 12353 (this is NBER macro annual 2006)

Cooley, T. and Dwyer, M. (1998), "Business Cycle Analysis without much Theory: A Look at Structural VARs, *Journal of Econometrics*, 83, 57-88.

Corradi, V. and Swanson, N. (2007), "Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historical Data," *Journal of Econometrics*, 136, 699-723.

Dedola, L. and Neri, S. (2007): What does a technology shock do? A VAR analysis with model-based sign restrictions, *Journal of Monetary Economics*

Del Negro, M. and Schorfheide, F. (2004): " Priors from General Equilibrium Models for VARs", *International Economic Review*, 95, 643-673.

Del Negro, M. and Schorfheide, F (2006) "Monetary Policy Analysis with Potentially Misspecified models", NBER working paper 13099.

Del Negro, M, Schorfheide, F., Smets, F. And R. Wouters (2005) On the fit of New Keynesian models, *Journal of Business and Economic Statistics.*

Erceg, C., D. Henderson, and A. Levin (2000): "Optimal Monetary Policy with Staggered Wage and Price Contracts," *Journal of Monetary Economics*, 46, 281–313.

Faust, J. and Leeper, E. (1997) Do Long Run Restrictions Really Identify Anything?, *Journal of Business and Economic Statistics*, 15, 345-353.

Fernandez-Villaverde, J., Rubio-Ramirez, J. , Sargent, T. and Watson, M. (2007) The ABC (and D) of structural VARs, *American Economic Review*, 97, 1021-1026.

Fisher, J. (2006), "The dynamic effect of Neutral and Investment specific technology Shocks ", *Journal of Political Economy*, , 114, 413-451.

Fry, R. and Pagan, A. (2007). "Some Issues in Using Sign Restrictions for Identifying Structural VARs", NCER Working Paper No 14.

Gali, J., and Rabanal, P (2004), "Technology shocks and Aggregate Fluctuations: How well does the RBC model fit post war US data?", NBER macro annual

Gali, J., and López-Salido, J. D. and Vallés, J. (2007), "Understanding the Effects of Government Spending on Consumption ", Journal of the European Economic Association, 5, 227-270'

Hansen, L. and Sargent, T., (1991), "Two Difficulties in Interpreting Vector Autoregressions", in Hansen, L. and Sargent, T., eds., *Rational Expectations Econometrics*, Westview Press: Boulder & London.

Kydland, F. and Prescott , E., 1996, "The Computational Experiment", *Journal of Economic Perspectives*, 10 (1)

Kim, J., 2000, " Constructing and Testing realistic Optimizing models of Monetary Policy", *Journal of Monetary Economics*, 45, 329-359.

Kilian, L. (1999) "Small Sample confidence Intervals for Impulse Response Functions", *Review of Economics and Statistics*, 218-230.

Iskrev, N (2007) How much do we learn from the estimation of DSGE models? A case study of Identification Issues in DSGE models, University of Michigan, manuscript.

Ireland, P. (2004) "Technology shocks in the New Keynesian Model", *Review of Economics and Statistics*, 86, 923-936.

Lippi, M. and Reichlin, L., (1994), "VAR Analysis, Non-Fundamental Representation, Blaschke Matrices", *Journal of Econometrics*, 63, 307- 325.

McGrattan, E (2004) Comment to Gali and Rabanal in the NBER macro annual

Moon, K. and Schorfheide, F. (2007) A Bayesian Look at Partially Identified Models, University of Pennsylvania, manuscript.

Pappa, E. (2005): "New Keynesian or RBC Transmission? The Effects of Fiscal Shocks in Labour Markets," CEPR Discussion Paper No 5313.

Paustian, M (2007) Assessing Sign Restrictions, *B.E. Journals of Macroeconomics,Topics*, 7(1), article 23.

Perotti, R. (2007) "In Search of a Transmission Mechanism of Fiscal Policy", forthcoming, *NBER Macro Annual*.

Rabanal, P., and J. Rubio-Ramirez (2005): "Comparing New Keynesian models of the business cycle: A Bayesian approach," *Journal of Monetary Economics*, 52, 1151-1166.

Ravenna, F. (2007): "Vector Autoregressions and Reduced Form Representations of Dynamic Stochastic General Equilibrium Models", *Journal of Monetary Economics*, 54, 2048-2064.

Rotemberg, J. and Woodford, M. (1997) , An Optimization based Econometric Framework for the evaluation of Monetary Policy, *NBER Macroeconomic Annual*, 12, 297-346.

Schorfheide, F. (2000), "Loss function based evaluation of DSGE models", *Journal of Applied Econometrics*, 15, 645-670.

Smets, F. and Wouters, R. (2003) An estimated Dynamic Stochastic General Equilibrium models of the Euro area, *Journal of the European Economic Association*, I, 1123-1175.

Uhlig, H. (2005), "What are the Effects of Monetary Policy? Results from an agnostic Identification procedure", *Journal of Monetary Economics*, 52, 381-419.

Appendix

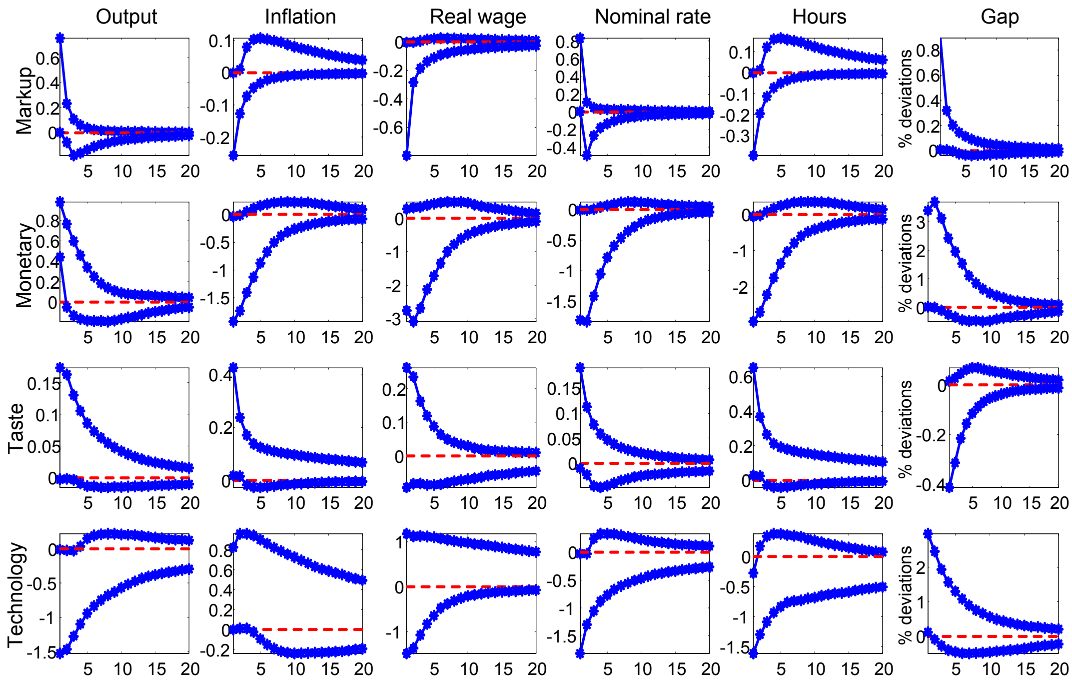


Figure A1: 95 percent response bands in the general model

	VAR1	VAR2	VAR1	VAR2	VAR1	VAR2	VAR1	VAR2
T=80	0.95	0.01	2.09	1.69	0.30	0.01	-1.60	-0.00
	-0.09	0.77	-0.91	-0.75	-0.19	-0.04	0.69	0.00
	-1.25	0.17	-5.97	-5.50	-1.04	-0.03	5.19	0.02
	1.29	-0.17	7.66	6.18	1.02	0.03	-5.87	-0.02
T=160	0.98	0.11	2.09	1.74	0.22	-0.06	-1.65	-0.00
	-0.05	0.75	-0.92	-0.77	-0.12	0.03	0.73	0.00
	-0.93	-0.19	-5.96	-5.64	-0.77	0.23	5.35	0.02
	0.94	0.19	7.67	6.36	0.77	-0.23	-6.07	-0.02
T=500	1.08	0.02	2.03	1.65	0.01	0.02	-1.64	0.00
	-0.10	0.80	-0.89	-0.73	-0.01	-0.01	0.72	-0.00
	-1.29	0.11	-5.70	-5.28	0.00	-0.06	5.27	-0.00
	1.32	-0.12	7.44	6.02	-0.01	0.06	-6.01	0.00
True	1.09	0.03	2.01	1.63	0.00	0.00	-1.63	0.00
	-0.12	0.81	-0.88	-0.72	0.00	-0.00	0.72	0.00
	-1.24	0.08	-5.63	-5.23	0.00	-0.00	5.23	0.00
	1.26	-0.09	7.38	5.98	0.00	0.00	-5.98	0.00

Table A.1: Estimated (averaged) and true VAR coefficients.

T=80	0.67	-0.29	-2.40	2.44
	-0.29	0.14	1.05	-1.07
	-2.40	1.05	8.66	-8.81
	2.44	-1.07	-8.81	8.96
T=160	0.69	-0.30	-2.49	2.53
	-0.30	0.15	1.08	-1.10
	-2.49	1.08	8.98	-9.13
	2.53	-1.10	-9.13	9.29
T=500	0.70	-0.30	-2.51	2.55
	-0.30	0.15	1.09	-1.11
	-2.51	1.09	9.04	-9.20
	2.55	-1.11	-9.20	9.36
True	0.70	-0.30	-2.53	2.57
	-0.30	0.15	1.10	-1.12
	-2.53	1.10	9.11	-9.27
	2.57	-1.12	-9.27	9.43

Table A.2 Estimated (average) and true covariance matrix.