

# MEASUREMENT WITH SOME THEORY: A NEW APPROACH TO EVALUATE BUSINESS CYCLE MODELS \*

Fabio Canova<sup>†</sup> and Matthias Paustian<sup>‡</sup>

First draft: November 2007;

This draft February 11, 2010

## Abstract

We propose a method to evaluate cyclical models which does not require knowledge of the DGP and the exact empirical specification of the aggregate decision rules. We derive robust restrictions in a class of models; use some to identify structural shocks and others to evaluate the model or contrast sub-models. The approach has good size and excellent power properties, even in small samples. We show how to examine the validity of a class of models, sort out the relevance of certain frictions, evaluate the importance of an added feature, and indirectly estimate structural parameters.

**JEL classification:** E32, C32.

**Keywords:** Misspecification, Sign restrictions, Shock identification, Model validation.

---

\*We thank Heather Anderson, Luca Benati, Adrian Pagan, Frank Schorfheide for comments and the participants of seminars at CREI, ECB, University of Zurich, Bank of Hungary, the conferences "How much structure in empirical models", Barcelona, and "Monetary policy in open economies", Sydney for suggestions. This paper replaces an earlier manuscript by the first author circulated with the title "Validating Monetary DSGE's with VARs". Canova acknowledges the financial support of the Spanish Ministry of Education grant ECO2009-08556 (Specification and Estimation of Models for Policy Analysis and Forecasting) and of the Barcelona Graduate School of Economics. The views expressed in this paper are those of the authors and not necessarily those of the Bank of England.

<sup>†</sup>ICREA-UPF, CREI, AMeN, CEPR; Department of Economics, UPF, Ramon Trias Fargas 25-27, 08005, Barcelona (Spain), Fabio.Canova@upf.edu

<sup>‡</sup>Bank of England, Threadneedle Street, London EC2R 8AH, United Kingdom, matthias.paustian@bankofengland.co.uk.

## 1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are nowadays regarded as the benchmark business cycles models for policy analysis and forecasting, both in academic and policy institutions. Their popularity stems from the attractive theoretical aspects and the good empirical performance they display, and from the useful forecasting properties they possess in the medium run, in particular, relative to single equation structural models or multiple equations time series specifications.

Existing business cycle models are, however, not problem free. Theoretically, many important features are modelled as black-box mechanisms; ad-hoc frictions are routinely added to match important aspects of the data; relevant real world phenomena have no role in standard constructions; crucial properties are often derived without any reference to parameter or model uncertainty. Empirically, the problems are numerous and varied. For classical estimation, a model is required to be the data generating process (DGP), up to a set of serially uncorrelated measurement errors. Since current business cycle models, even in the large scale versions used in policy institutions, fail to meet such a requirement, it is difficult to credibly entertain structurally estimated models. This assumption is unnecessary to derive meaningful posterior distributions. Still, even in a Bayesian framework, it is hard to interpret misspecified estimates, unless an explicit loss function is considered (see Schorfheide (2000)). In any case, the inherent misspecification models possess causes the likelihood function to be poorly behaved, making numerical difficulties widespread and Bayesian estimation strongly dependent on the prior selection. On the other hand, abundant identification problems (see Canova and Sala (2009) and Canova and Gambetti (forthcoming)) and the severe mismatch between theoretical and empirical concepts of business cycles (see Canova (2009)) render structural estimation and policy conclusions generically whimsical. The empirical validation of business cycle models is also difficult: models impose fragile restrictions on the magnitude of interesting statistics and evaluation techniques for misspecified, hard to identify models are underdeveloped. Those based on classical asymptotic ideas are unsuited and, if we exclude a few notable cases (Schorfheide and Del Negro (2004) and (2009)), those currently available are computationally intensive, only permit evaluation relative to a benchmark model, are ill suited when both models have low posterior probability and focus on statistical fit, rather than the relevance of the economic discrepancy. For useful policy experiments, meaningful welfare calculations, and informative conditional forecasting exercises researchers sorely need techniques which are simple, reproducible, effective in measuring economic discrepancy and informative about the reasons for its existence.

This paper presents a methodology to validate classes of business cycle models and to select sub-models in a class. It employs the flexibility of SVAR techniques against model misspecification, the insights of computational experiments (see e.g. Kydland and Prescott (1996)) and pseudo-Bayesian predictive analysis (see e.g. Canova (1995)) to design probabilistic measures of fit which

are informative about the economic relevance of a class of models and can discriminate among locally alternative DGPs. We take seriously the objection that existing business cycle models are at best approximations of a portion of the DGP. We are sympathetic to the claim that too little predictive analysis is typically performed prior to structural estimation and that existing models may be unsuited for traditional estimation and testing because of specification problems and identification failures. We also pay attention to the fact that the quantitative restrictions are typically fragile and design criteria which employ robust qualitative theoretical implications.

The analysis starts from a class of models which has an approximate state space representation once (log-)linearized around the steady state. We examine the dynamics of the endogenous variables in response to the disturbances for alternative members of the class using a variety of parameterizations. While magnitude restrictions depend on specification details, the sign of the impact responses and, at times, the dynamic shape of the responses are much more robust to parameter and specification uncertainty. We use a subset of theoretically robust restrictions to identify structural disturbances and use the responses of unrestricted variables to construct qualitative and quantitative measures of economic discrepancy between the class and the data or to select a member within the class. The approach is constructive: if the discrepancy is deemed large at any stage of the evaluation process, the class of models can be respecified and the analysis repeated.

Our methodology has a number of advantages. First, it does not require the true DGP to be a member of the class of models we consider: we only need that the subset of the robust qualitative restrictions employed for identification has a counterpart in the data. If this were not the case, our approach can detect the problem and allow to immediately stop the evaluation process. Second, our approach does not need the probabilistic structure to be fully specified in order to be operative. Thus, ad-hoc dynamic additions as well as shock proliferation become fully dispensable (see Kocherlakota (2007) for a related argument) Third, by focusing shock identification and model testing on robust model-based qualitative restrictions, our methodology catches several birds with one stone: it de-emphasizes the quest for a good calibration; it gives economic content to identification restrictions used in SVARs analyses; it shields researchers against specification problems. These aspects of the methodology should be attractive for applied researchers struggling to evaluate potentially misspecified models delivering quantitatively weak testable restrictions. Fourth, the approach can be used in a limited information or full information mode, and has degrees of freedom that can be used to make shock identification and model testing stronger. Fifth, the procedure requires small computing power, it is easily reproducible and applicable to several interesting classes of models.

We show that the approach can recover the sign of the impact response of unrestricted variables to the identified shocks, measure interesting qualitative features of the DGP, and exclude potentially relevant candidate DGPs with high probability for a variety of structural designs, even when sample uncertainty exists. Moreover, because the evaluation procedure focuses on qualitative

rather than quantitative implications of the theory, it delivers reasonable conclusions, even when the empirical model is incorrectly specified relative to the DGP. Finally, since the emphasis is on robust restrictions, we can distinguish sub-models in situations where standard approaches fail.

We illustrate with two examples how the methodology can be used to sort the frictions consistent with the transmission mechanism we observe in the data; to analyze the general relevance of a class of models; to evaluate the importance of a feature added to an existing class of models; and to get "estimates" of parameters which are typically non-identifiable when aggregate data and standard econometric techniques are used. We show, in the class of models popularized by Christiano et al. (2005) and Smets and Wouters (2003), that the impact response of the real wage to government spending shocks can be used to discriminate price from wage rigidities in the data and the dynamic responses of hours to various types of technology shocks employed to evaluate the general quality of the fit. We demonstrate that price frictions may not be crucial to characterize cyclical fluctuations in the US and raise doubts about the quality of the approximation provided by this class of models for the data. We also show, in the class of models with a portion rule-of-thumb agents suggested by Gali et al. (2007), that the presence of a large number of non-optimizing consumers is insufficient to make consumption responses to government spending shocks positive and indicate how the robust restrictions of the class can be employed to measure the sign, the magnitude and the shape of consumption responses in the data. Since the share of non-optimizing agents needed to quantitatively match the conditional consumption dynamics is unrealistically large, the validity of this class of models for policy and interpretation exercises is also seriously called into question.

The rest of the paper is organized as follows. Section 2 describes the methodology; section 3 studies the properties of the procedure in a series of controlled experiments. Section 4 evaluates two standard business cycle models. Section 5 concludes.

## 2 A sign restriction approach to evaluation

It is our presumption that current business cycle models, while useful to qualitatively characterize conditional dynamics, are still too stylized and feature too many black-box frictions to be taken seriously, even as an approximation to part of the DGP of the actual data (a point made, with different emphasis, also by Chari et al. (2009)). Since this misspecification will not necessarily vanish completing the probabilistic space of the model with measurement errors, ad-hoc shocks or artificial dynamics, we do not follow the standard approach of finding parameters that make the augmented model and the data quantitative "close" and statistically measure the magnitude of the discrepancy. Instead, we derive a set of dynamic implications, which are qualitatively robust to the parameterization and to the specification of the model within a class; use some of these implications to recover structural disturbances in the data and employ others to measure the quality of the

model's approximation to the data or to select competing sub-models in the class.

To describe our approach we need some notation. Let  $F(w_t^s(\theta), \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M}) \equiv F^s(\theta)$  be a set of continuous model-based functions, which can be computed conditional on the structural disturbances  $\epsilon_t$ , using models in the class  $\mathcal{M}$ .  $F^s(\theta)$  could include impulse responses, conditional cross correlations, distributions of conditional turning points, etc., and depends on the model-produced series  $w_t^s(\theta)$ , where  $\theta$  are the structural parameters, and, possibly, on the parameters of their VAR representation, 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 data-based functions, conditional on the reduced form shocks  $u_t$  and the parameters of the VAR representation of the data. We assume that  $w_t$  is a  $q \times 1$  vector. We take the class  $\mathcal{M}$  to be broad enough to include sub-models with interesting economic features.  $\mathcal{M}$  could be, e.g., one of the New Keynesian models used in the literature and the sub-models versions where wage stickiness or price indexation are shut off. The class  $\mathcal{M}$  is misspecified in the sense that even if there exists a  $\theta_0$  such that  $\alpha_0 = \alpha_0(\theta_0)$  or  $\alpha_1 = \alpha_1(\theta_0)$  or both,  $F(w_t^s(\theta), \alpha_0(\theta_0), \alpha_1(\theta_0)|\epsilon_t, \mathcal{M}) \neq F(w_t, \alpha_0, \alpha_1|u_t)$ .

Among all possible  $F^s(\theta)$  functions, we restrict attention to the subset  $\tilde{F}^s(\theta)$  which are robust: the  $J_1 \times 1$  vector  $\tilde{F}_1^s(\theta) \subset \tilde{F}^s(\theta)$  is used for shock identification and the  $J_2 \times 1$  vector  $\tilde{F}_2^s(\theta) \subset \tilde{F}^s(\theta)$  for evaluation purposes,  $\tilde{F}_1^s(\theta) \neq \tilde{F}_2^s(\theta)$ .  $\tilde{F}^s(\theta)$  is termed robust if either  $sgn(F^s(\theta_1)) = sgn(F^s(\theta_2))$ , or if  $sgn(F^s(\theta_1)|\mathcal{M}_j) = sgn(F^s(\theta_2)|\mathcal{M}_j)$ ,  $\forall \theta_1, \theta_2 \in [\theta_l, \theta_u]$ , where  $sgn$  is the sign of  $F^s$ ;  $\theta_l, \theta_u$  are the upper and lower range of economically reasonable parameter values and  $\mathcal{M}_j \in \mathcal{M}$ . Thus,  $\tilde{F}_1^s(\theta)$  contains functions whose sign is independent of the sub-model and the parameterization;  $\tilde{F}_2^s(\theta)$  contains functions whose sign is independent either of the sub-model and the parameterization (if the generic fit is evaluated) or of the parameterization (if sub-models are compared). The economic question to be investigated dictates what  $\tilde{F}_1^s(\theta)$  and  $\tilde{F}_2^s(\theta)$  will be.

## 2.1 The algorithm

The evaluation procedure involves five steps:

- 1) Find robust implications of the class  $\mathcal{M}$ . That is, find  $\tilde{F}^s(\theta)$  and select  $\tilde{F}_1^s(\theta)$  and  $\tilde{F}_2^s(\theta)$ .
- 2) Use  $\tilde{F}_1^s(\theta)$  to identify disturbances in the data. That is, find the set of  $\alpha_0$  that minimizes  $\mathcal{I}_{[sgnF_1(w_t^k, \alpha_0, \alpha_1|u_t) - sgnF_1(w_t^{k,s}, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M}) \neq 0]}$ , subject to  $A_0 A_0' = \Sigma_u$ ,  $\alpha_0 = A_0 H$ ,  $H H' = I$  where  $\theta \in [\theta_l, \theta_u]$ ,  $\mathcal{I}_{[\cdot]}$  is a counting measure,  $\Sigma_u$  the covariance matrix of reduced form disturbances,  $k = 1, 2, \dots, q_1 < q$ . If there is no  $\alpha_0$  such that  $0 \leq \mathcal{I}_{[\cdot]} \leq \iota$ , some  $\iota \geq 0$ , choose another set of  $\tilde{F}_1^s(\theta)$  and, if none remains, stop the evaluation process.
- 3) Evaluate the performance of the class **qualitatively** computing (a)  $S_1^k(\mathcal{M}) = \frac{100}{N} \times \mathcal{I}_{[sgn\tilde{F}_2(w_t^k, \hat{\alpha}_0, \alpha_1|u_t) - sgn\tilde{F}_2(w_t^{k,s}, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M}) = 0]}$  and /or (b)  $S_2^k(\mathcal{M}) = \frac{100}{N} \times$

$\mathcal{I}_{[shp\tilde{F}_2(w_t^k, \hat{\alpha}_0, \alpha_1|u_t) - shp\tilde{F}_2^s(w_t^{k,s}, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M})]_{=0}}$ , where *shp* is the dynamic shape of  $\tilde{F}_2$ ,  $\hat{\alpha}_0$  are the  $N$  values of  $\alpha$  obtained in step 2),  $\tilde{F}_2^s$  hold for interesting members of  $\mathcal{M}$ ,  $k = q_1 + 1, \dots, q$ .

- 4) If the performance in 3) is satisfactory, **quantitatively** evaluate  $\mathcal{M}$ . For example, compute  $Pr(\tilde{F}_2^s(\theta) \leq \tilde{F}_2(\hat{\alpha}_0)) \forall \theta \in [\theta_l, \theta_u]$  or the degree of overlap between  $D(\tilde{F}_2^s(\theta))$  and  $D(\tilde{F}_2(\alpha_0))$ , where the distributions  $D$  are obtained randomizing over  $\theta$  and the  $\alpha_0$  found in 2)
- 5) Evaluate sub-models of the class. For example, for  $h = 1, 2$  compute  $S_3^k(\mathcal{M}_h) = \frac{100}{N} \times \mathcal{I}_{[sgn\tilde{F}_2(w_t^k, \hat{\alpha}_0, \alpha_1|u_t) - sgn\tilde{F}_2^s(w_t^{k,s}, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M}_h)]_{=0}}$  or  $S_4^k(\mathcal{M}_h) = \frac{100}{N} \times \mathcal{I}_{[shp\tilde{F}_2(w_t^k, \hat{\alpha}_0, \alpha_1|u_t) - shp\tilde{F}_2^s(w_t^{k,s}, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, \mathcal{M}_h)]_{=0}}$  and choose the  $\mathcal{M}_h$  that minimizes  $S(\mathcal{M}_h) = \sum_{j=1}^{J_2} w_j^1 S_{3j}(\mathcal{M}_h) + \sum_{j=1}^{J_2} w_j^2 S_{4j}(\mathcal{M}_h)$ , where  $\sum_j w_j^1 + \sum_j w_j^2 = 1$  are chosen weights.
- 6) If the performance in 3) or 4) is unsatisfactory, choose another class of models or add to  $\mathcal{M}$  features that may help to reduce the discrepancy. Otherwise, undertake policy analyses (welfare analyses, conditional forecasting exercises, etc.) as needed.

In the first step 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 inflation in response to monetary policy shocks is unchanged when we vary the structural parameters within a reasonable range, and this is true for an interesting subset of models in  $\mathcal{M}$ , we call this a robust implication. Robustness is not generic: many features are sensitive to the parameterization and to the type of frictions present in the model. Moreover, since models are misspecified, magnitude restrictions are unlikely to hold. Hence, the robust implications we consider take the form of sign restrictions on conditional responses, primarily in the impact period. We focus on conditional functions as they are more informative than unconditional ones about the features of the class  $\mathcal{M}$ .

In the second step, we make the class of models and the data share qualitative aspects of their conditional functions. This step is easily implementable using the approaches of Canova and De Nicolò (2002), Uhlig (2005) or Rubio-Ramirez et al. (forthcoming). One can "weakly" or "strongly" identify disturbances, by imposing a small or a large number of robust restrictions, across shocks and/or variables. Since standard rank and order conditions are not applicable, how minimal this set of restrictions should be in general, is discussed below. Contrary to traditional practices, we derive identification restrictions explicitly from a class of models and employ only robust qualitative constraints. This way, we construct conditional dynamics without conditioning on any particular member of the class nor on its parameterization. When the chosen set of restrictions fails to hold in the data, one would either impose an alternative set of robust restrictions, or, if all are exhausted and no disturbances with the required properties found, go back to the drawing board and select a different class.

The third step is similar to the one employed in computational experiments where some moments are used to calibrate the structural parameters; others to check the goodness of the theory. Here robust sign restrictions are employed to identify structural disturbances; the sign and the shape of the dynamic responses of unrestricted variables can be used to check the quality of the model's approximation to the data. We differ from standard practices because, at both stages, we only consider robust qualitative implications and because evaluation is probabilistic.

When the analysis requires quantitative answers to certain questions, conditional forecasting exercises or welfare calculations, the quality of the class can be further assessed via Monte Carlo methods, i.e. using measures of distance between distributions of outcomes (as e.g. Canova (1995)). The computational costs of this step are minimal since distributions of outcomes in theory are obtained in step 1), and distributions of data outputs in step 2). Quantitative evaluation is not a substitute for a qualitative one: classes of models can be easily eliminated and the burden of evaluation reduced if a qualitative check is performed first. Clearly, if the quantitative performance is unsatisfactory, the selected class should not be used for policy exercises (see also Del Negro and Schorfheide (2009)).

Researchers are often concerned with the relative likelihood of sub-models in a class differing in terms of microfoundations, frictions, or functional forms. One can compare sub-models using qualitative devices such as the sign and shape of selected responses to shocks. For example, two sub-models in a class may produce different sign for the response of hours to technology shocks. Once restrictions which are common to the two sub-models are used to identify technological disturbances, the response of hours to these shocks could be used to discriminate sub-models. If sub-models differ in a number of implications, a weighted average of counting measures can be used to select the model with the smaller discrepancy with the data. The weights could be optimally chosen, but we prefer to let them be free parameters - depending on the problem, one may want to weight different functions differently. If robustness is a concern, pseudo-bayesian averaging, where a scaled version of  $S_j^k(\mathcal{M}_h)$ ,  $j = 3, 4$  is employed as weight, can be used. Candidate sub-models could be nested and or non-nested: our method works in both setups. However, in the latter case we need to assume that agents know the economy they live in and only the applied investigator face model uncertainty.

## 2.2 Discussion

The procedure is informative about the properties of a class of models and the dimensions of mismatch with the data. For example, shape differences may suggest what type of amplification mechanism may be missing and sign differences the frictions/shocks that need to be introduced. Note that conditional dynamics can be analyzed in response to one or several shocks at a time.

The approach we propose compares favorably to existing approaches for at least two reasons.

Classical estimation and inference are asymptotically justified under the assumption that the model is the DGP and Bayesian inference is problematic without a loss function. In addition, both classical and Bayesian estimation have hard time to deal with the population identification problems highlighted e.g. in Canova and Sala (2009) making testing difficult (see Del Negro and Schorfheide (2008)). Both issues are relatively minor in our setup: the use of robust identification restrictions shields researchers from model and parameter misspecification; since the mapping between structural parameters and the coefficients of the decision rule is not used, lack of parameters identification is less of a problem. Since the set of  $\alpha_0$ 's in step 2) is not necessarily a singleton, the procedure recognizes that the relationship between the  $\alpha_i, i = 0, 1$  and the  $\theta$ s may not be unique.

SVAR analyses are often criticized because shock identification is not linked to the theory that it is used to interpret the results (see e.g. Canova and Pina (2005)). Since we employ theory based robust sign restrictions, such a problem is absent here. A number of authors (see Christiano, et. al (2006), Fernandez-Villaverde et. al. (2007), Ravenna (2007)), and Chari et. al (2008)) have indicated that another form of subtle misspecification may be present in SVARs. While the literature has cast this problem into an "invertibility" issue, it is best to think of it as an omitted variable problem. Let the decision rules of a log-linearized model be:

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

where  $e_t \sim iid(0, \Sigma_e)$ ,  $x_{1t}$  are the states,  $x_{2t}$  the controls,  $e_t$  the innovations in the disturbances and  $A(\theta), B(\theta), C(\theta), D(\theta)$  continuous differentiable functions of the structural parameters  $\theta$ . Thus, the log-linearized decision rules are members of a class of VAR(1) models of the form:

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

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  $x_{1t}$  and  $x_{2t}$ . Then, the representation for  $y_{2t}$  is

$$(I - F_{22}\ell - F_{21}F_{12}(1 - F_{11}\ell)^{-1}\ell^2)y_{2t} = [G_2 - (F_{21}(1 - F_{11}\ell)^{-1}G_1)\ell]e_t \tag{3}$$

While the model for  $y_{2t}$  is an ARMA( $\infty, \infty$ ), the impact effect of the shocks in (2) and (3) is identical, both in terms of magnitude and sign. Thus, as long as robust sign restrictions are imposed on impact, this form of misspecification does not affect shock identification <sup>1</sup>.

---

<sup>1</sup>In small samples, estimates of  $G_2$  will be biased making standard magnitude restrictions unlikely to hold. As we show later, when  $G_2$  is not close to zero, sign restrictions will hold even in small samples.



### 2.3 Comparing our approach to the literature

The methodology we propose is related to early work by Canova, Finn and Pagan (1994); and to the recent strand of literature identifying VAR disturbances using sign restrictions (see Canova and De Nicolò (2002) or Uhlig (2005)). It is also related to Del Negro and Schorfheide (2004) and (2009), and Del Negro et. al. (2006) who use the data generated by a cyclical model as a prior for reduced form VARs. Two main differences set our approach apart: we condition the analysis on a general class of models rather than a single one; we only work with qualitative restrictions rather than quantitative ones. This focus allows generic forms of model misspecification to be present and vastly extends the range of structures for which model evaluation becomes possible.

Corradi and Swanson (2007) developed 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 economic reasons for the discrepancy between the model and the data. Fukac and Pagan (2008) also suggest using limited information methods to evaluate business cycle models but consider quantitative restrictions on single equations of the model while we focus on qualitative implications induced by certain disturbances. 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 of fit for model comparison exercises.

## 3 The evaluation procedure in a controlled experiment

To examine the properties of our procedure in realistic setups, we consider a class of New-Keynesian models without capital, employed e.g. by Rabanal and Rubio Ramirez (2005) among others, which allows for habit in consumption, and for price and wage rigidities. We choose this class because several sub-models of interest are nested into the general setup and the structure is flexible and tractable. In the first part, we investigate the properties of our procedure in population. Later on, we discuss whether sampling and specification uncertainty make a difference.

### 3.1 The class of models

The equilibrium conditions, with variables expressed in log-deviations from the steady state, are

$$\lambda_t = E_t \lambda_{t+1} + (R_t - E_t \pi_{t+1}) \quad (4)$$

$$\lambda_t = e_t^b - \frac{\sigma_c}{1-h} (y_t - h y_{t-1}) \quad (5)$$

$$y_t = e_t^z + (1-\alpha)n_t \quad (6)$$

$$mc_t = w_t + n_t - y_t \quad (7)$$

$$mrs_t = -\lambda_t + \sigma_l n_t \quad (8)$$

$$w_t = w_{t-1} + \pi_t^w - \pi_t \quad (9)$$

$$\pi_t^w - \mu_w \pi_{t-1} = \kappa_w [mrs_t - w_t] + \beta(E_t \pi_{t+1}^w - \mu_w \pi_t) \quad (10)$$

$$\pi_t - \mu_p \pi_{t-1} = \kappa_p [mc_t + e_t^\mu] + \beta(E_t \pi_{t+1} - \mu_p \pi_t) \quad (11)$$

$$R_t = \rho_R R_{t-1} + (1 - \rho_R) [\gamma_\pi \pi_t + \gamma_y y_t] + e_t^R \quad (12)$$

Equation (4) is the consumption Euler equation:  $\lambda_t$  is the marginal utility of consumption,  $R_t$  the nominal interest rate,  $\pi_t$  price inflation. Equation (5) defines the marginal utility of consumption with external habit formation and  $e_t^b$  is a preference shock. The production function is in (6);  $e_t^z$  is a productivity disturbance and  $n_t$  are hours worked. Real marginal costs  $mc_t$  are defined in (7), and  $w_t$  is the real wage. Equation (8) gives an expression for the marginal rate of substitution,  $mrs_t$ . Equation (9) is an identity linking real wage growth to the difference between nominal wage and price inflation. The wage and price Phillips curves arising from Calvo nominal rigidities are in (10) and (11).  $\mu_p$  ( $\mu_w$ ) parameterizes the degree of backward-lookingness in price (wage) setting;  $e_t^\mu$  is a price markup shock, and  $\pi_t^w$  wage inflation. The slopes of the curves are  $\kappa_p \equiv \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p} \frac{1-\alpha}{(1-\alpha+\alpha\epsilon)}$  and  $\kappa_w \equiv \frac{(1-\zeta_w)(1-\beta\zeta_w)}{\zeta_w(1+\varphi\sigma_l)}$ , respectively. The central bank adjusts the nominal interest rate  $R_t$  according to the rule in (12). The four disturbances ( $e_t^z, e_t^b, e_t^R, e_t^\mu$ ) are driven by mutually uncorrelated, mean zero innovations. The productivity shock  $e_t^z$  and the preference shock  $e_t^b$  have autocorrelation coefficients  $\rho_z$  and  $\rho_b$ , respectively. The monetary shock  $e_t^R$  and the markup shock  $e_t^\mu$  are iid. The standard deviations of the innovations are  $(\sigma^z, \sigma^b, \sigma^R, \sigma^\mu)$ .

At least five sub-models of interest are nested into this general structure, which we label M - a flexible price, sticky wage model ( $\zeta_p = 0$ ), which we label M1; a sticky price, flexible wage model ( $\zeta_w = 0$ ), which we label M2; a flexible price and flexible wage model ( $\zeta_p = 0, \zeta_w = 0$ ), which we label M3; a model with no habits ( $h = 0$ ), which we label M4, a model with no indexation ( $\mu_p = 0, \mu_w = 0$ ), which we label M5 - and this allows us to conduct meaningful testing exercises.

To find sign restrictions that hold across parameter values and for sub-models in the class, we specify for each parameter a uniform distribution over an interval, chosen to be large enough to include theoretically reasonable values, existing structural estimates or values used in calibration exercises - see third column of Table 1. For example, the interval for the risk aversion coefficient contains the values used in the calibration literature (typically 1 or 2) and the higher values typically employed in the asset pricing literature (see e.g. Bansal and Yaron (2004)), while the intervals for the habit and the Calvo parameters include, roughly, the universe of possible values considered in the literature. Since the discount factor  $\beta$  and the markup parameters  $\epsilon$  and  $\varphi$  are not separately identified - they enter the two Phillips curves as composites, together with the price and wage stickiness parameters - they are fixed in our exercises.

We draw a large number of parameter vectors, compute impulse responses for each draw and, with the collection of responses, construct pointwise 90 percent response intervals. Table 2 reports

Parameter	Description	Support	DGP1	DGP2
$\beta$	discount factor	0.99	0.99	0.99
$\epsilon$	elasticity in goods bundler	6	6	6
$\varphi$	elasticity in labor bundler	6	6	6
$\sigma_c$	risk aversion coefficient	[1.00, 5.00]	2.00	2.00
$\sigma_l$	inverse Frish elasticity of labor supply	[0.00, 5.00]	1.74	1.74
$h$	habit parameter	[0.00, 0.95]	0	0
$\zeta_p$	probability of keeping prices fixed	[0.00, 0.90]	0	0.75
$\zeta_w$	probability of keeping wages fixed	[0.00, 0.90]	0.62	0
$\mu_p$	indexation in price setting	[0.00, 0.80]	0	0
$\mu_w$	indexation in wage setting	[0.00, 0.80]	0	0
$\alpha$	1 - labor share in production function	[0.30, 0.40]	0.36	0.36
$\rho_r$	inertia in Taylor rule	[0.25, 0.95]	0.74	0.74
$\gamma_y$	response to output in Taylor rule	[0.00, 0.50]	0.26	0.26
$\gamma_\pi$	response to inflation in Taylor rule	[1.05, 2.50]	1.08	1.08
$\rho_z$	persistence of productivity	[0.50, 0.99]	0.74	0.74
$\rho_b$	persistence in taste process	[0.00, 0.99]	0.82	0.82
$\sigma_z$	standard deviation of productivity		0.0388	0.0388
$\sigma_\mu$	standard deviation of markup		0.0316	0.0316
$\sigma_b$	standard deviation of preferences		0.1188	0.1188
$\sigma_r$	standard deviation of monetary		0.0033	0.0033
$\sigma_m$	standard deviation of measurement error		0.0010	0.0010

Table 1: Supports for the parameters and DGPs used in the experiments.

the sign of the interval in the impact period. For each shock, there is a column for the general model and one for each sub-model; a '+' indicates robustly positive responses; a '-' robustly negative responses; a '?' responses which are not robust; and 'NA' responses which are zero by construction. Figure A.1 in the appendix shows the range of dynamic outcomes. The observable variables we consider are the nominal rate ( $R_t$ ), the real wage ( $w_t$ ), the inflation rate ( $\pi_t$ ), output ( $y_t$ ), and hours ( $n_t$ ).

Many of the impact responses have robust signs, both across parameterizations and sub-models. For example, positive markup shocks increase production costs for any sub-model and parameterization; thus, for a given demand, production, the real wage and employment contract while inflation and the nominal rate increase. In general, taste shocks are the disturbances delivering less robust impact responses across sub-models and the real wage the variable whose impact response is less robust within sub-models. The sign of the real wage responses crucially depends on the relative degree of wage and price stickiness; given the ranges we employ, it is then natural that the real wage may fall or rise in response to several shocks.

The impact response of the real wage to monetary disturbances is of particular interest since it differs in sign for sub-models in the class featuring different frictions. In sub-model M2 (sticky

	Markup shocks					Monetary shocks						
	M	M1	M2	M3	M4	M5	M	M1	M2	M3	M4	M5
$R_t$	+	+	+	+	+	+	+	+	+	+	+	+
$w_t$	-	-	-	-	-	-	?	+	-	NA	?	?
$\pi_t$	+	+	+	+	+	+	-	-	-	-	-	-
$y_t$	-	-	-	-	-	-	-	-	-	NA	-	-
$n_t$	-	-	-	-	-	-	-	-	-	NA	-	-

	Taste shocks					Technology shocks						
	M	M1	M2	M3	M4	M5	M	M1	M2	M3	M4	M5
$R_t$	+	+	+	?	+	+	-	-	-	-	-	-
$w_t$	?	-	?	-	?	?	?	+	?	+	?	?
$\pi_t$	+	+	?	?	+	+	-	-	-	-	-	-
$y_t$	+	+	+	+	+	+	+	+	+	+	+	+
$n_t$	+	+	+	+	+	+	-	-	-	-	-	-

Table 2: Signs of the impact response intervals to shocks, different models. A '+' indicates robustly positive responses; a '-' robustly negative responses; a '?' responses which are not robust; and 'NA' responses which are zero by construction. M is the general model, in M1  $\zeta_p = 0$ ; in M2  $\zeta_w = 0$ ; in M3  $\zeta_p = 0$  and  $\zeta_w = 0$ ; in M4  $h = 0$ ; in M5  $\mu_p = 0$  and  $\mu_w = 0$ .

prices, flexible wages) real wages fall. With flexible wages, workers are on their labor supply schedule and, on impact,  $w_t = \left( \frac{\sigma_c}{1-h} + \frac{\sigma_l}{1-\alpha} \right) y_t$ , so that real wages negatively comove with monetary shocks. In sub-model M1 (flexible prices and sticky wages), workers are off their labor supply schedule and from the firm's labor demand schedule,  $w_t = -\frac{\alpha}{1-\alpha} y_t$  - real wages positively comove with monetary shocks. Thus, while it is difficult to distinguish between sticky price and sticky wage models using unconditional measures of wage cyclicality, sign restrictions on the impact response of the real wage to monetary shocks can separate the two sub-models in the class.

### 3.2 Population analysis

We pick the flexible price, sticky wage sub-model M1 as our DGP. The parameters used in simulating the "pseudo-actual" data are the fourth column of table 1 and are close to those estimated by Rabanal and Rubio-Ramirez (2005). We endow the researcher with (4)-(12) and its solution and let both the model dynamics and the covariance matrix of the reduced form errors  $\Sigma$  be known. We ask whether the responses of the real wage to various shocks can be recovered with high probability employing different subsets of the robust restrictions present in table 2 in alternative VAR systems and identifying shocks either jointly or separately.

We estimate the matrix of impact coefficients as follows: i) we draw a large number of normal,

zero mean, unitary matrices; ii) employ the QR decomposition and construct impact responses as  $\alpha_0 = S * Q$ , where  $SS' = \Sigma$ ; and iii) keep the responses satisfying the restrictions we impose. To make sure results are stable, we draw until 10000 candidates satisfying the restrictions are found.

### 3.2.1 Can we recover the true model?

The empirical model is composed of 5 variables: the nominal rate, output, inflation, hours and the real wage. Since the economy features 4 structural shocks, we sidestep singularity issues by attaching a measurement error to the law of motion of the real wage. We identify disturbances (a) jointly, using robust impact restrictions on all variables but the real wage; (b) jointly, using robust impact restrictions on all variables but hours and the real wage; (c) individually, the markup shock; (d) individually, the monetary shock. In (c) and (d) we use robust impact restrictions on all variables but the real wage. In addition to the basic DGP, we also examine two alternative setups, one where the standard deviation of monetary shocks is 10 times larger and one where the standard deviation of the markup shocks is 10 times larger, and for each setup we repeat the four experiments. Table 3 reports the percentage of correctly signed impact real wage responses.

Identified shocks	Basic				Larger monetary shocks				Larger markup shocks			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
Markup	99.8		99.8		99.9		99.9		100		100	
Monetary	75.7	76.2		74.9	92.4	90.1		89.7	58.2	59.0		50.6
Taste	98.8	98.3			99.2	99.3			97.8	95.8		
Technology	99.7				99.7				96.2			
Supply		99.7				99.1				99.9		

Table 3: Percentage of cases where the impact real wage response is correctly signed. The VAR includes output, real wages, hours, inflation and the nominal rate. In case (a) output, inflation, nominal rate and hours are restricted and shocks are jointly identified; in case (b) output, nominal rate and inflation are restricted and a supply shock, a monetary and a markup shock are identified; in cases (c) and (d) output, inflation, nominal rate and hours are restricted and a markup or a monetary shock are separately identified. In panels two and three the standard deviation of either the monetary or of the markup shocks is set 10 times larger.

Our procedure recognizes the qualitative features of the DGP with high probability, when the ideal conditions we consider hold. Two features of table 3 deserve some comments. First, the number of shocks identified and, to a less extent, the number of identification restrictions employed matter for the results. The larger imprecision obtained when only monetary shocks are identified, (compare cases (a) and (d) in each panel) comes from the fact that monetary shocks are contaminated - they pick up features of other structural disturbances and of the measurement error.

Second, as in Paustian (2007), the relative strength of the shock signal matters. For example, the real wage effects of markup and taste shocks are easy to measure because their signal is relatively strong, making conclusions largely independent of the number of restrictions used and the number of shocks identified. For monetary shocks, whose standard deviation is relatively small, having a sufficient number of restrictions is important to capture the sign of the real wage impact response. When the standard deviation of monetary shocks is increased, the precision with which our approach recognizes the sign of the impact effect uniformly increases, highlighting the trade-off between the strength of the signal and number of identification restrictions used.

Studies of the transmission of monetary shocks are abundant in the last 15 years and several researchers have used sign restrictions to identify these disturbances in the data. Since such disturbances are likely to have small relative variability, our results indicate that their transmission could be mismeasured, unless a sufficiently large number of restrictions is employed. In general, since the relative volatility of many structural shocks is unknown, being too agnostic in the identification process may have important costs for inference.

In the appendix, we demonstrate that these conclusions hold true when hours is dropped from the empirical model. A 4 variable VAR is fundamentally different from a 5 variable VAR since, in the latter, a state variable is missing - the observed real wage is a contaminated signal of the true one. Ravenna (2007) and Chari et. al. (2008) indicated that such an omission may be dangerous for inference if standard structural VARs are estimated. Our results confirm the point made in section 2: when robust sign restrictions on the impact response of certain variables are used for identification purposes, misspecification of the VAR is less crucial for inference.

### 3.2.2 Can we exclude alternative models?

A sticky price, flexible wage sub-model and a flexible price, sticky wage sub-model are local to each other as far as the sign of impact responses is concerned. As table 2 shows, the impact effect of a number of variables to the four shocks in the two sub-models is similar. The procedure can recover the sign of the real wage response to monetary shocks well when the flexible price, sticky wage sub-model M1 is the DGP. Would the answer be different if the sticky price, flexible wage sub-model M2 and the parameterization listed in the last column of table 1 characterizes our DGP? In other words, can we exclude with high probability that sub-model M1 is the DGP just by looking at the sign of the impact responses of the real wage to monetary shocks?

The answer is positive. In the three experiments considered (identifying all shocks using the impact restrictions on output, inflation, hours and the nominal rate; identifying monetary, taste and supply shocks using impact restrictions on output, inflation and the nominal rate; and identifying only monetary shocks) the percentage of incorrectly recognized cases ranges between 0.4 and 1.3 percent. Could this conclusion be due to the selection of the parameters of the DGP?

To examine this possibility, we have considered two other sets of experiments. First, we have increased the standard deviation of either the monetary shocks or the markup by a factor of ten. The conclusion are broadly unchanged: the fraction of impact real wage responses to monetary shocks that is incorrectly signed never exceeds 8.0 percent. Second, we have allowed the parameters to be randomly and uniformly drawn from the intervals in table 1 - in this case, we draw 200 parameter vectors, setting  $\theta_w = 0$  for every draw, and for each vector, we draw 10000 identification matrices. When only monetary shocks are identified, the sign of the impact real wage response is incorrectly identified, on average, 3.21 percent of the times - the numerical standard error is 5.47. Thus, the exact parameterization has little influence on the results.

Why is our procedure successful in capturing the DGP and in excluding local DGPs as potential data generators? The answer is simple. While the range of impact real wage responses to monetary shocks generated randomizing the parameters of the DGP in M1 and M2 is relatively large, the degree of overlap of the distribution of responses is minimal. Thus, we can tell apart the two sub-models with high probability because the theory has sharp and alternative implications for the real wage responses to monetary shocks. The answer would be different if the theoretical implications of different sub-models are more muddled. For example, the response of the real wage to technology shocks in M2 is not robust and the percentage of incorrect cases exceeds 25 percent under some identification configuration. Thus, only robust restrictions should be used for testing purposes.

These results are interesting also from a different perspective. Canova and Sala (2009) and Iskrev (2007) have shown that classical econometric approaches have difficulties in separating sticky price and sticky wage models, because the distance function constructed using dynamic responses or the likelihood function are flat in the parameters controlling price and wage stickiness. Similar results are reported by Del Negro and Schorfheide (2008), when Bayesian methods are used. Our semi-parametric approach, which exploits the idea that only robust restrictions should be used for identification and evaluation, can instead give sharp answers to this question.

### 3.2.3 Summarizing the shape of the dynamic responses

The evaluation analysis has so far concentrated on the sign of the impact response of a variable left unrestricted in the identification process. For many empirical purposes this focus is sufficient: business cycle theories do not typically have robust implications for the magnitude or the persistence of the responses to shocks. At times, however, the shape of the dynamic responses is of interest and one may want to know either in which percentile of the estimated distribution of responses the true responses lie after, say, one year or whether there exists a location measure that reasonably approximates, say, certain conditional multipliers.

Figure 1 plots the median of the set of identified real wage responses to shocks, horizon by horizon, and the true real wage responses in the Basic setup, case (a) of table 3. The median is a

reasonable measure of the impact response of real wages to all shocks, both in a qualitative and in a quantitative sense, while it is an imperfect estimator of the true conditional real wage dynamics, at least as far as the responses to monetary shocks are concerned. Nevertheless, in terms of the sign of the responses, the median is an acceptable summary measure. Relative to other location measures, it is slightly better than the average response and very similar to the trimmed mean (computed dropping the top and the bottom 25 percent of the responses).

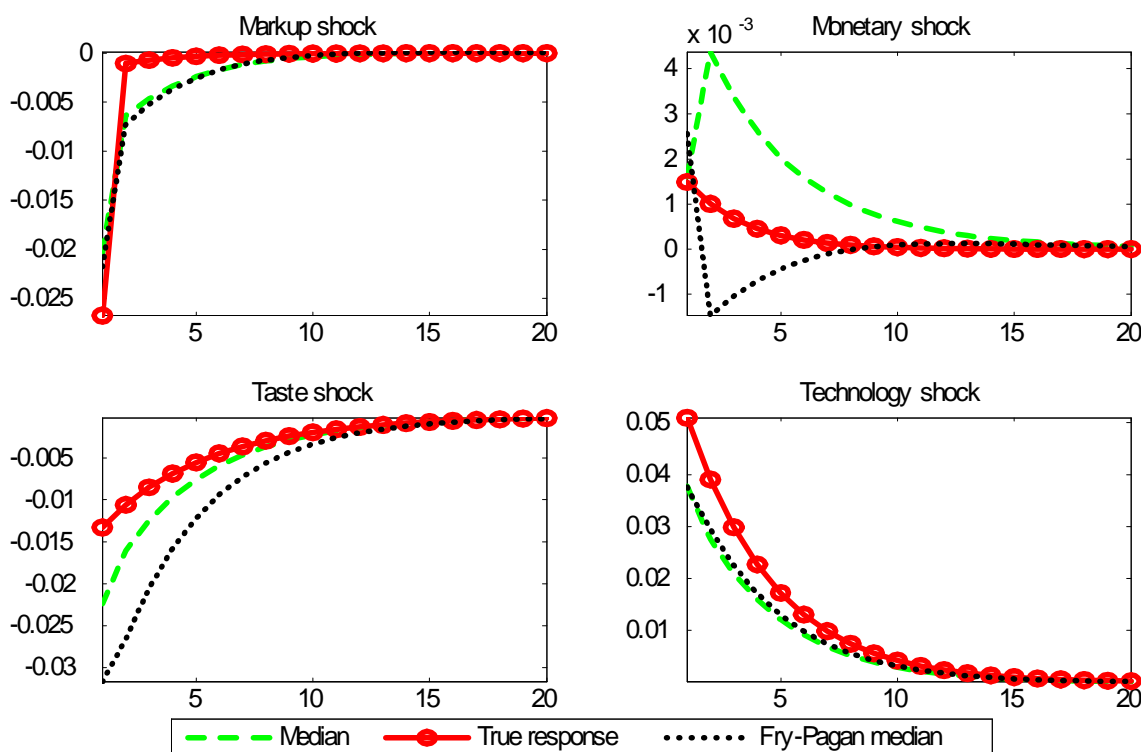


Figure 1: Real wage responses to shocks.

Fry and Pagan (2007) have criticized the practice of using the median of the distribution of responses as location measure 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. As a consequence, structural shocks may not be uncorrelated, and structural analyses difficult to interpret. As an alternative, they suggest to use the single identification matrix that comes closest to producing the median impulse response for all variables. In our exercises, the correlation among identified shocks, computed using the median, is indeed significant and ranges from 0.59 to 0.89



in absolute value. Therefore, the concern of Fry and Pagan seems legitimate. However, as figure 1 shows, Fry and Pagan’s median is not a uniformly superior summary measure: it is similar to our median measure for markup and technology shocks; it is quantitatively worse for taste shocks; and for monetary shocks, it produces real wage responses with the wrong sign after a few horizons. Thus, if attention focuses on the responses to monetary shocks, it is unclear which measure should be used and, if the sign of the dynamics is of interest, having uncorrelated shocks may be worse.

To know more about the performance of the two summary location measures, we have calculated the contemporaneous correlation between the true disturbances and the estimated disturbances computed using the Fry and Pagan median and between the true disturbances and the estimated disturbances obtained by taking the median value of the identification matrix (i.e. the matrix which produces on impact the median response).

	Median Identification				Fry and Pagan Median			
	Markup	Monetary	Taste	Technology	Markup	Monetary	Taste	Technology
Markup	0.94	-0.24	0.17	-0.15	0.35	0.01	-0.53	-0.38
Monetary	-0.42	0.81	-0.36	0.11	-0.27	0.88	-0.27	0.18
Taste	0.76	-0.02	0.62	0.14	-0.24	0.17	0.72	-0.41
Technology	0.74	-0.15	0.33	0.54	0.11	0.34	0.10	0.79

Table 4: Correlations between true shocks (columns) and estimated shocks (rows). The DGP is the sticky wage, flexible price model M1; the VAR includes output, real wages, hours, inflation and the nominal rate; shocks are identified using impact restrictions on output, inflation, hours and nominal rate. In the first panel the median rotation matrix is used.

Table 4 summarizes these correlations. Both measures appear to produce qualitatively similar results: estimated shocks are mostly correlated with the true shocks of the same type, but some contamination is present. As far as the correlation between identified monetary shocks with true shocks (second column in each panel) distortions appear to occur because, when the median identification matrix is used, estimated monetary shocks are negatively correlated with true markup shocks and, when the Fry and Pagan median is used, estimated monetary shocks are positively correlated with technology shocks. Thus, while theoretically appealing, the Fry and Pagan median is not clearly preferable to other summary statistics reflecting identification uncertainty.

We have conducted a number of additional exercises to check whether the performance of location statistics is affected by changes in the experimental design. The results agree with what we had in the previous subsections: identifying more shocks or increasing the strength of the variance signal improves the dynamic performance of the median; the dimensionality of the VAR has little influence on the dynamic properties of the median; the exact nature of the DGP makes

no difference for the conclusions we reach.

### 3.3 Does sampling and specification uncertainty matter?

The ideal conditions considered so far are useful to understand the properties of the procedure but unlikely to hold in practice. Do conclusions change if the autoregressive parameters and the covariance matrix of the shocks are estimated prior to the identification of the structural disturbances?

To capture estimation uncertainty, we consider 200 replications of each experiment we have run. In each replication, we simulate data, keeping the parameters fixed and injecting in the decision rules random noise (and measurement error) in the form of normal iid shocks with zero mean and standard deviations as reported in table 1. We consider samples with 80, 160 and 500 points - 20, 40 and 125 years of quarterly data. For each replication, we estimate a fixed finite order BVAR, where a close to non-informative conjugate Normal-Wishart prior is used. We prefer the option of an arbitrary lag length because it is the one typical used in practice even though, for our DGP, it introduces an additional source of misspecification - the decision rules imply that a VAR( $\infty$ ) should be used. Later, we examine what happens if the lag length is optimally selected. We jointly draw from the posterior of the parameters, the covariance matrix of the shocks and the identification matrices until 2000 draws satisfying the restrictions are found. We compute pointwise medians and pointwise credible 90 percent posterior intervals for 20 horizons for the response of the real wage to monetary shocks at each replication. We summarize the results by computing the median (or the average) value across replications of the median estimate and the interval containing 90 percent of the estimated 90 percent intervals at each horizon.

We generate data from a sticky wage, flexible price model with one measurement error. A BVAR with the nominal rate, output, inflation, hours, and the real wage is estimated. We identify shocks imposing sign restrictions on the impact responses of the nominal rate, output, inflation and hours. Figure 2 reports the real wage responses to monetary shocks for different sample sizes and for different lag lengths, when only monetary shocks are identified. The corresponding figure when all shocks are identified is in the appendix.

Three features of figure 2 stand out. First, sample uncertainty is small relative to identification uncertainty: the magnitude of the intervals decreases as the sample size increases for each lag length, but the differences between T=80 and T=500 are small. In standard SVARs, biases in the estimates of the dynamics are usually of an order of magnitude larger than those in the estimated covariance matrix. This is true also here. However, biases in estimated VAR coefficients are also relatively small. In fact, the estimated median dynamics have similar properties as the sample size increases, and even with T=80 the median and the true dynamics are reasonably close. Thus, even a loose Bayesian prior, is effective in eliminating most of the problems noted by Kilian (1999). Second, the envelope of the bands is wide and includes the zero line at every horizon. Thus, literally speaking,

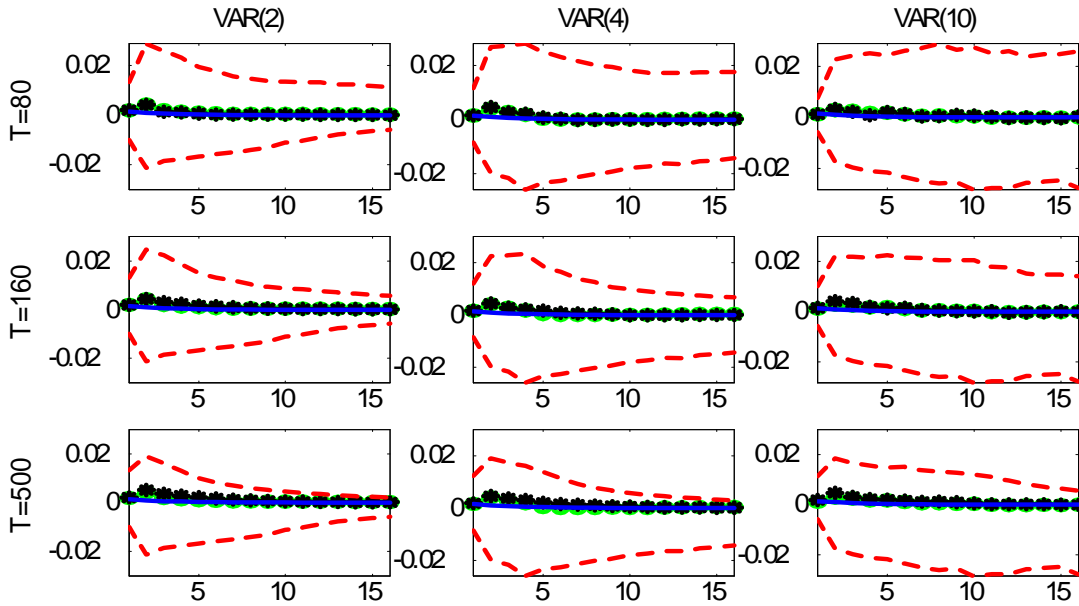


Figure 2: Real wage responses to monetary shocks, Monte Carlo results.

it is difficult to statistically pin down the sign of real wage responses. One could make estimation results look better, by changing the uncertainty measures, for example, reporting the median (or the average) of the upper and lower 90 percent credible intervals across replications. Large bands are the results of the considerable identification uncertainty intrinsic to the approach. Below, we describe a way to reduce it. Third, changing the lag length of the VAR has little consequences on the outcomes. With a larger number of lags, the bands become generally larger, especially when  $T=80$ , but the dynamics of the median are unchanged. As it is shown in the appendix, this remains true also when the lag length is selected with BIC. Hence, none of the problems highlighted by Chari, et al. (2008) appear to be present here. Finally, the number of shocks we identify has little consequences on the quality of the outcomes. As expected, identifying more shocks makes the bands smaller for any  $T$  and any VAR length, but only marginally so, and the properties of the median are unchanged <sup>2</sup>.

---

<sup>2</sup>We have also computed coverage rates - that is, the probability that the true response falls within the estimated credible interval - but we have decided to omit them since they provide little new information. Coverage rates for partially identified models are generally smaller than those computed with classical methods and a standard nominal size because of the way identification uncertainty is treated in the two contexts (see Moon and Schorfheide (2009)). In general, the coverage rate for the wage in response to monetary shocks is 60% on impact and increases to about 96%

All the conclusions we have obtained in population are unaffected by the presence of sample and specification uncertainty. For example, we can still recognize the DGP and exclude local sub-models with high probability looking at the impact response of the real wage to monetary shocks (see table A2 in the appendix) and the performance of the median as summary measure is unaffected by sample and, more remarkably, by specification uncertainty. The appendix also reports summary tables for experiments conducted when the sticky wage, flexible price model M1 is the DGP and the standard deviation of markup shocks or the standard deviation of the monetary shocks and when the sticky price; when flexible wage model M2 is the DGP and either four or five variables are used in the VAR with different lags. All our basic results hold in these alternative setups.

In sum, sample and specification uncertainty have minor consequences on the performance of our approach. Sample uncertainty is small relative to identification uncertainty (see Kilian and Murphy, 2009, for related evidence); specification uncertainty is a minor problem for our methodology.

### 3.4 Advice to the users

Our procedure has good size and power properties. However, three main ingredients are needed to give the methodology its best chance to succeed. First, it is important not to be too agnostic in the identification process. Sign restrictions are weak and this makes identification uncertainty important (see Manski and Nagy (1998) for a similar result in micro settings). Thus, it is generally easier to recognize the DGP when more variables are restricted, for a given number of identified shocks, or more shocks are identified. Since, as we have mentioned, theoretical sign restrictions at horizons larger than the impact one are often whimsical, constraints on the dynamic responses should be avoided at the identification stage.

Our experiments also showed that credible intervals tend to be large. Since the methodology delivers partially identified models, expecting the same degree of estimation precision as for exactly identified models is foolish. However, probabilistic summary statistics are informative about the features of the DGP, even when asymptotically-based standard normal tests are not. If one insists on using these tests, one should be aware that they are meaningful only if a sufficient number of restrictions are employed (the amplitude of the bands is inversely proportional to the number of identification restrictions employed), and that smaller confidence intervals (say, 68 percent or interquartile ranges) are more appropriate when identification uncertainty is large.

Second, estimation biases should be, when possible, reduced since they may compound with identification uncertainty. In the experiments we have run, estimation biases were small, even in small samples, but this needs not to be the case for every possible design. A close to non-informative

---

at longer horizons for the basic VAR(2). As the sample size increases, coverage is slightly lower since the estimated bands shrink but the change is small. Coverage rates are not much affected by the VAR length: with a VAR(10), the coverage rate for the impact wage response to monetary shocks is 62%.

prior was sufficient to reduce them. Other approaches, see Kilian (1999), may work as well.

Third, inference is very reliable when the analysis focuses on the dynamics induced by shocks with a stronger relative variance signal. However, even when the shock signal is weak, as the monetary shocks in our designs, systematic mistakes are absent. Thus, even though a-priori is hard to say which shock dominates (structural estimation does not help since the magnitude of the relative variances depends on the degree of misspecification present in various equations), and pathological examples can always be constructed (see Paustian (2007) or Fry and Pagan (2007)), relative differences in the variance signal become a problem only in very extreme circumstances. When interesting shocks are suspected to generate a weak relative signal, we recommend users to employ plenty of identification restrictions and to consider a class of models with a sufficiently rich shock structure. These two conditions were sufficient to insure a good performance in all experiments we run.

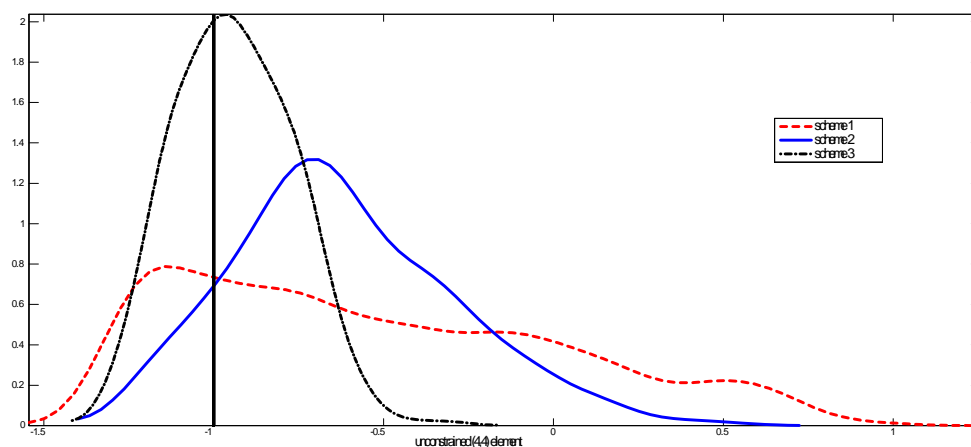


Figure 3: Response under different identification schemes. Scheme 1 sign restrictions, one shock; Scheme 2 sign plus uniqueness restrictions, on shock; Scheme 3 sign restrictions all shocks. Vertical bar: true value.

In theory, it is often the case that disturbances generate a unique pattern of impact responses for the endogenous variables. However, in practice, responses are not restricted to satisfy this uniqueness condition. Thus, especially when a small subset of the shocks is identified, it is possible that shocks disregarded in the analysis generate similar pattern of responses. This multiplicity has no reason to exist and may make inference weaker than it should. But, in practice, what do we lose by failing to impose the uniqueness condition in identification? Typically a lot.

To show this, we generate density estimates of the unconstrained (4, 4) element of the matrix

$$D = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$$

in a static four variable VAR,  $y = De$ , where  $e$  has diagonal variance with elements  $[1,1,1,2]$ , identifying the last shock only using restrictions on the  $(j, 4) > 0, j = 1, 2, 3$  elements of the matrix (scheme 1), identifying the last shock using the same restrictions and the restriction that the other three shocks can not generate a similar pattern of responses (scheme 2) and identifying all the shocks using the restrictions on the  $(j, i), j = 1, 2, 3; i = 1, \dots, 4$  elements of the matrix. Figure 3 shows that the distribution of responses in scheme 1 (dotted line) and in scheme 2 (solid line) looks very different: 30 percent of the mass of the estimated distribution is above zero in scheme 1 and only 9 percent is above zero when the additional uniqueness restrictions are imposed; the median of the distribution is a better estimator of the true value in scheme 2. Thus, while not a substitute for identifying all the shocks, which can be seen gives very precise information about the sign and the magnitude of the unrestricted element, imposing the uniqueness condition may help to sharpen inference when only a subset of the shocks is identified.

## 4 Examples

This section presents two examples illustrating how the methodology can be used to sort out the frictions consistent with the transmission mechanism observed in the data; to analyze the relevance of a class of models; to evaluate the importance of an added feature; and to "estimate" parameters which are typically non-identifiable with aggregate data and standard techniques.

### 4.1 Evaluating a benchmark specification

#### 4.1.1 The class of models

The class of models we consider is regarded as the benchmark for policy analysis and forecasting in the literature (see Christiano, et. al. (2005) and Smets and Wouters (2003)). The structure features nominal frictions (sticky nominal wage and price setting, backward wage and inflation indexation), real frictions (habit formation in consumption, investment adjustment costs, variable capital utilization and fixed costs in production) and nests several specifications one maybe interested in analyzing. The class has three blocks and its log-linearized representation (around the

steady state) is as follows. The aggregate demand block is:

$$y_t = c_y c_t + i_y i_t + g_y e_t^g \quad (13)$$

$$c_t = \frac{h}{1+h} c_{t-1} + \frac{1}{1+h} E_t c_{t+1} - \frac{1-h}{(1+h)\sigma_c} (R_t - E_t \pi_{t+1}) + \frac{1-h}{(1+h)\sigma_c} (e_t^b - E_t e_{t+1}^b) \quad (14)$$

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} E_t i_{t+1} + \frac{\phi}{1+\beta} q_t - \frac{\beta E_t e_{t+1}^I - e_t^I}{1+\beta} \quad (15)$$

$$q_t = \beta(1-\delta) E_t q_{t+1} - (R_t - E_t \pi_{t+1}) + (1-\beta(1-\delta)) E_t r_{t+1} \quad (16)$$

Equation (13) 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 (14) is a dynamic IS curve:  $e_t^b$  is a preference shock,  $\sigma_c$  the coefficient of relative risk aversion and  $h$  the coefficient of external habit formation. The dynamics of investment are in equation (15);  $\phi$  represents the elasticity of the costs of adjusting investments,  $q_t$  the value of existing capital,  $e_t^I$  a shock to the investment's adjustment cost function and  $\beta$  the discount factor. Equation (16) characterizes Tobin's  $q$ : the current value of the capital stock positively depends on its expected future value and its expected return, and negatively on the ex-ante real interest rate. The aggregate supply block is:

$$y_t = \omega(\alpha k_{t-1} + \alpha \psi r_t + (1-\alpha)n_t + e_t^z) \quad (17)$$

$$k_t = (1-\delta)k_{t-1} + \delta i_t \quad (18)$$

$$\pi_t = \frac{\beta}{1+\beta\mu_p} E_t \pi_{t+1} + \frac{\mu_p}{1+\beta\mu_p} \pi_{t-1} + \kappa_p m c_t \quad (19)$$

$$w_t = \frac{\beta}{1+\beta} E_t w_{t+1} + \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta} E_t \pi_{t+1} - \frac{1+\beta\mu_w}{1+\beta} \pi_t + \frac{\mu_w}{1+\beta} \pi_{t-1} - \kappa_w \mu_t^W \quad (20)$$

$$n_t = -w_t + (1+\psi)r_t^k + k_{t-1} \quad (21)$$

Equation (17) is the aggregate production function. In equilibrium  $\psi r_t$  equals the capital utilization rate and  $e_t^z$  is a total factor productivity (TFP) shock. Fixed costs of production are given by  $\omega - 1$  and  $\alpha$  is the capital share. The law of motion of capital accumulation is in equation (18). Equation (19) links inflation to marginal costs,  $m c_t = \alpha r_t^k + (1-\alpha)w_t - e_t^z + e_t^{\mu_p}$  and  $e_t^{\mu_p}$  is a markup shock. 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  $\zeta_p$ , the probability that firms face for not being able to change prices in the Calvo setting. The parameter  $\mu_p$  determines the degree of price indexation. Equation (20) links the real wage to expected and past wages, to inflation and to the marginal rate of substitution between consumption and leisure,  $\mu_t^W = w_t - \sigma_l n_t - \frac{\sigma_c}{1-h}(c_t - h c_{t-1}) - e_t^{\mu_w}$ , where  $\sigma_l$  is the inverse of the elasticity of hours to the

real wage,  $e_t^{\mu_w}$  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 (21) follows from the equalization of marginal costs. The monetary rule is

$$R_t = \rho_R R_{t-1} + (1 - \rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^R \quad (22)$$

where  $\varepsilon_t^R$  is a monetary policy shock.

Equations (13) to (22) define a system of 10 equations in ten unknowns,  $(\pi_t, y_t, c_t, i_t, q_t, l_t, w_t, k_t, r_t, R_t)$ . Given these variables, the productivity-wage gap ( $gap_t = \frac{y_t}{n_t} - w_t$ ) can be easily generated. The model features seven exogenous disturbances: TFP,  $e_t^z$ , investment-specific,  $e_t^I$ , preference,  $e_t^b$ , government spending,  $e_t^g$ , monetary policy,  $e_t^R$ , price markup  $e_t^{\mu_p}$  and labor supply,  $e_t^{\mu_w}$  shocks. The vector of disturbances  $S_t = [e_t^z, e_t^I, e_t^b, e_t^g, e_t^R, e_t^{\mu_p}, e_t^{\mu_w}]'$ , satisfies:

$$\log(S_t) = (I - \boldsymbol{\varrho}) \log(\bar{S}) + \boldsymbol{\varrho} \log(S_{t-1}) + V_t \quad (23)$$

where  $V \sim iid(0', \Sigma_v)$ ,  $\boldsymbol{\varrho}$  is diagonal with roots less than one in absolute value and  $\bar{S} = E(S)$ .

Parameter	Description	Support
$\sigma_c$	risk aversion coefficient	[1,6]
$\sigma_l$	inverse Frish labor supply elasticity	[0.5,4.0]
$h$	consumption habit	[0.1,0.8]
$\omega$	fixed cost	[1.0,1.80]
$1/\phi$	adjustment cost parameter	[0.0001,0.01]
$\delta$	capital depreciation rate	[0.015,0.03]
$\alpha$	capital share	[0.15,0.35]
$1/\psi$	capacity utilization elasticity	[0.1,0.6]
$\zeta_p$	degree of price stickiness	[0.4,0.9]
$\mu_p$	price indexation	[0.2,0.8]
$\zeta_w$	degree of wage stickiness	[0.4,0.9]
$\mu_w$	wage indexation	[0.2,0.8]
$\varepsilon^w$	steady state markup in labor market	[0.1,1.8]
$g_y$	share of government consumption	[0.10,0.25]
$\rho_R$	lagged interest rate coefficient	[0.2,0.95]
$\gamma_\pi$	inflation coefficient on interest rate rule	[1.1,3.0]
$\gamma_y$	output coefficient on interest rate rule	[0.0,1.0]
$\boldsymbol{\varrho}_i$	persistence of shocks $i = z, b, I, \mu_p, \mu_w$	[0,0.9]

Table 5: Supports for the structural parameters.

To find robust implications, we split the parameter vector  $\theta = (\theta_1, \theta_2)$ , where  $\theta_1 = (\beta = 0.99, \pi^{ss} = 1.016)$  are fixed parameters while  $\theta_2$  are parameters which are allowed to vary. Table 5



gives their intervals over which they can vary: the ranges are larger than the prior ranges considered, e.g., in the Bayesian estimation, and we are much more agnostic about the probability of certain parameter configurations than most papers in the literature.

#### 4.1.2 The identification restrictions

Table 6 reports the signs of the 68 percent impact responses intervals to the seven shocks, for a subset of the endogenous variables. As in section 3, a '+' indicates a robustly positive sign, a '-' a robustly negative sign and a '?' a sign which is not robust.

	TFP	Investment	Markup	Labor supply	Monetary	Taste	Government	Government No price rigidities	Government No wage rigidities
$\Delta y_t$	+	+	+	+	+	+	+	+	+
$\pi_t$	-	-	-	-	+	+	+	+	+
$\Delta c_t$	+	-	+	+	+	+	-	-	-
$\Delta gap_t$	+	?	-	+	-	-	-	-	-
$\Delta w_t$	+	-	+	-	+	+	?	-	+
$\Delta n_t$	-	+	+	+	+	+	+	+	+

Table 6: Sign of the impact response intervals to shocks.

The table provides useful identification information. Notice first, that TFP, investment specific, markup and labor supply disturbances increase output growth and decrease inflation on impact while the other three disturbances produce positive comovements of these two variables. Thus, shocks moving the aggregate supply and shocks moving the aggregate demand can be separated using these two variables. Second, government expenditure shocks can be distinguished from the other demand shocks since the impact response of consumption growth is negative with government expenditure disturbances and positive with the other two shocks. Third, there are enough (mutually exclusive) restrictions to separately identify the four supply disturbances. In fact, investment specific shocks make consumption growth fall on impact while the other three shocks induce a positive impact consumption growth response. Moreover, the impact response of the gap growth is positive in response to TFP and labor supply shocks and negative in response to markup shocks. Finally, real wage growth instantaneously falls in response to labor supply and investment specific shocks and increases in response to the other two shocks. Since these restrictions are valid in all the sub-models we have examined, e.g., with no habit, with full capacity utilization, with utility which is log in consumption or linear in leisure, with no wage stickiness or indexation, etc., they are representative of the class of models, qualify as robust, and can potentially be used to extract

the disturbances of interest in the data.

Table 6 displays two other interesting features. The impact response of real wage growth to a government expenditure shock depends on whether there are price or wage rigidities. Thus, it can be used to examine the relative importance of the two types of nominal rigidities in the data. In addition, hours growth robustly fall in response to TFP shocks and robustly increase in response to the other three technology shocks on impact. Since these restrictions hold in all the sub-models, except when investment adjustment costs are set to zero, they can be used as a general purpose statistic to evaluate the quality of the model's approximation to the data.

We use a VAR with 6 variables and two lags. We take US data for the sample 1948:2- 2007:2 from the FRED data bank at the Fed of St. Louis and estimate a BVAR using a diffuse prior. GDP growth is measured using per-capita real chain weighted GDP; inflation using the quarterly change of the GDP deflator; consumption growth using real per-capita private consumption expenditure; the gap growth using labor productivity growth (measured as output per worker) and the real wage compensation growth; wage growth using the compensations of employed in the business sector and hours growth using per-capita hours; again in the business sector. We jointly identify the four supply shocks and the government expenditure disturbance.

#### 4.1.3 Real wage responses to spending shocks

Government expenditure shocks are obtained restricting the sign of the impact response of output growth, inflation, hours growth, gap growth and consumption growth. We jointly draw from the distribution of BVAR parameters and orthonormal matrices until 1000 draws satisfying the restrictions are found. The first panel of figure 4 reports the median and the posterior 68 credible interval for the responses of real wage growth: the impact response is significantly negative - price frictions appear to be absent - but the response turns positive after one quarter.

The specification of the BVAR does not matter: eliminating the constant, using a more informative prior, or changing the lag length is irrelevant for the conclusions. For example, in the second panel of figure 4, we report the median and the posterior 68 credible interval in a 6 lags BVAR. On impact the interval is still entirely negative and, while the medium term dynamics are estimated with larger errors, the median response path in the first two panels has similar patterns.

Since the restrictions this class of models imposes on the dynamics of consumption growth are at odds with existing empirical evidence (see e.g. Perotti (2007) and the next example) we have also identified generic "demand" disturbances, without imposing restrictions on consumption growth. The dynamics of real wage growth are broadly unchanged (see third panel of figure 4): the impact real wage growth response to the identified demand shock is negative but since one less restrictions is used, response intervals at the first few horizons are wider.

It is well documented in the literature (see e.g. Schorfheide and Del Negro (2008) or Canova

and Sala (2009)) that standard estimation procedures have hard time to distinguish which nominal friction matters most with a finite sample of data. It is therefore remarkable that our procedure can tell them apart and that the answer is unambiguous: conditional on the class of models, price rigidities are unimportant.

To make the conclusion stronger, one would like to know how trustworthy the class of models is. Some authors, e.g. Smets and Wouters (2003),(2007), Christiano et al. (2005) claimed that the model fits the data well. Others, e.g. Del Negro et al. (2007) have raised important doubts. What does our procedure tell us when the response of hours growth to technological disturbances are used to measure the quality of the model's approximation to the data?

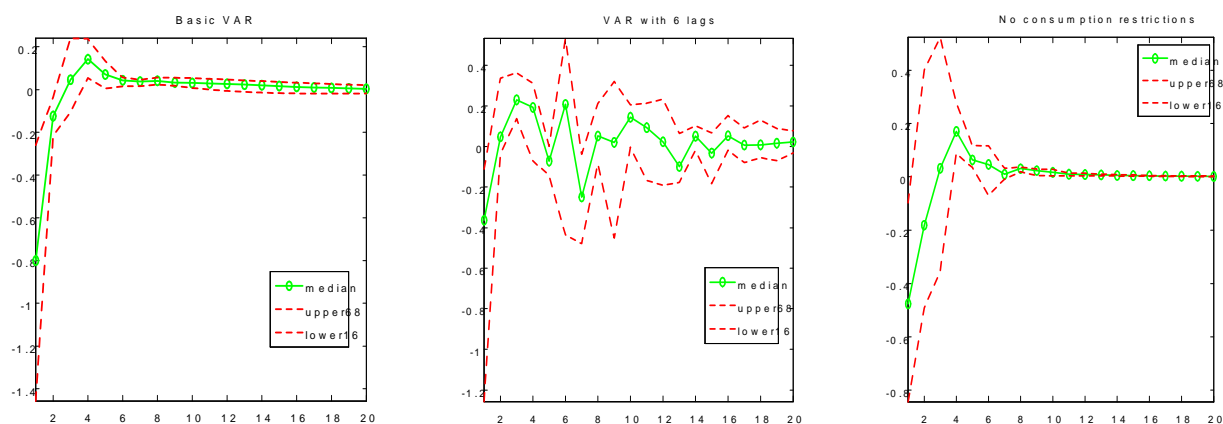


Figure 4: Responses of real wage growth to government expenditure shocks.

#### 4.1.4 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 been cast into a RBC vs. New-Keynesian microfoundations (see Rabanal and Gali (2004) and Chari et al. (2008)), researchers have started distinguishing various technology shocks (Fisher (2006)) and offer different explanations of the evidence. Rather than entering the controversy, this subsection asks whether the selected class of models produces hours dynamics in response to technology shocks which are consistent with the data?

Figure 5 reports, in the top panel, the 68 percent intervals of hours growth responses to the four shocks in the theory and, in the bottom panel, the median and the posterior 68 credible intervals for hours growth responses to the same four shocks in the data, when shocks are identified by restricting the sign of the impact response of output growth, inflation, consumption growth, hours growth and gap growth. Clearly, the model's approximation to the conditional dynamics of hours growth is limited. While the sign of the median hours growth response in the data is consistent with the theory (as in Paustian (2007)), the impact effect, which could robustly signed in theory

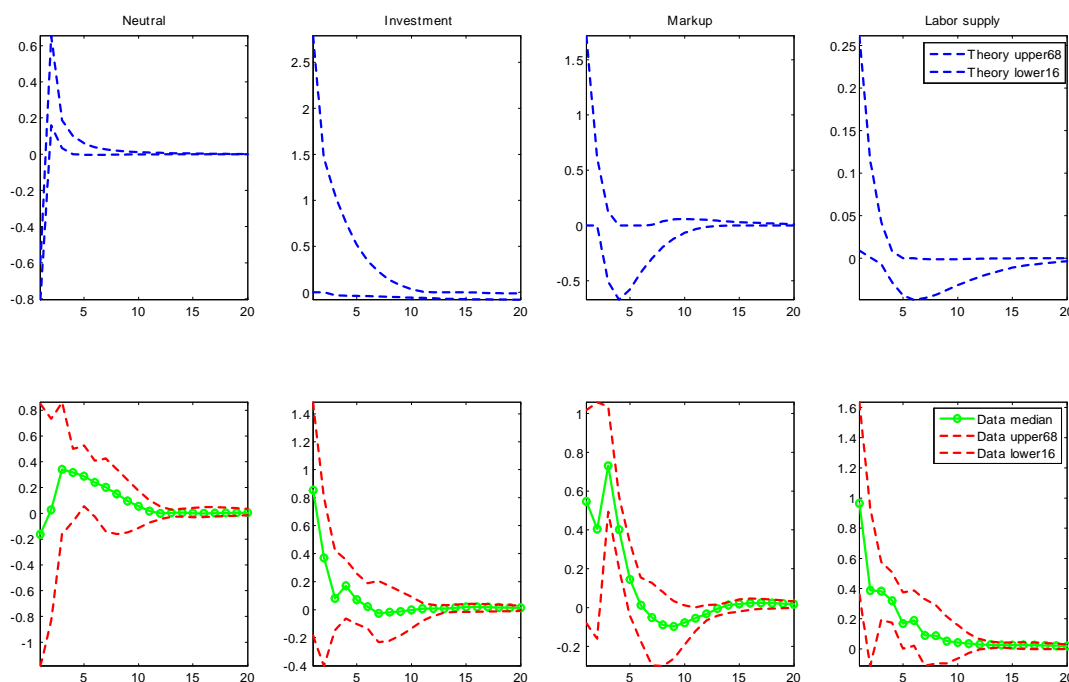


Figure 5: Responses of hours growth to technology shocks

for all shocks, can be signed with high probability only for labor supply shocks; the shape of 68 posterior interval of hours growth dynamics differs from 68 percent interval we had in theory; the magnitude of the responses is altered. In general, the best match is obtained with markup and labor supply shocks; the poorest with TFP and investment specific shocks.

In sum, while in the class of New-Keynesian models we consider the procedure unambiguously favours wage frictions to characterize the impact response of the real wage to government shocks, it also raises important doubts about the quality of the approximation provided by this class of models for the data. In particular, the robust restrictions the class possesses are insufficient to sign the impact responses of hours growth to several supply disturbances, to statistically characterize the dynamics responses and to quantitative evaluate the relative importance of various disturbances for hours growth fluctuations.

## 4.2 Adding a particular friction

It is well known that standard business cycle models have a hard time to produce the private consumption dynamics in response to government consumption expenditure shocks generated by structural VARs (see e.g. Perotti (2007)). However, one should also be aware that the restrictions used in the VARs are not explicitly derived from any theoretical specification proved to be consistent with the data. Gali et al. (2007) have taken a relative standard New Keynesian class of models and showed that adding one particular friction (a large portion of non-Ricardian consumers) can make the theory consistent with the existing structural VAR evidence. The question we investigate here is whether the sign of consumption responses in the data matches the one in theory, once the robust restrictions implicit in the class are used to identify government consumption shocks.

### 4.2.1 The class of models

The log-linearized conditions for the class of models we consider are

$$q_t = \beta E_t q_{t+1} + [1 - \beta(1 - \delta)] E_t r_{t+1}^k - (R_t - E_t \pi_{t+1}) + e_t^{rp} \quad (24)$$

$$i_t - k_{t-1} = \eta q_t \quad (25)$$

$$k_t = (1 - \delta)k_{t-1} + \delta i_t \quad (26)$$

$$c_t^o = c_{t+1}^o - (R_t - E_t \pi_{t+1}) \quad (27)$$

$$c_t^r = \frac{1 - \alpha}{\mu c_y} (w_t + n_t^r) - \frac{1}{c_y} t_t^r \quad (28)$$

$$w_t = c_t^j + \sigma_l n_t^j \quad j = o, r \quad (29)$$

$$r_t^k = m c_t + e_t^z + (1 - \alpha)(n_t - k_{t-1}) \quad (30)$$

$$w_t = m c_t + e_t^z - \alpha(n_t - k_{t-1}) \quad (31)$$

$$y_t = e_t^z + (1 - \alpha)n_t + \alpha k_{t-1} \quad (32)$$

$$y_t = c_y c_t + i_y i_t + g_y e_t^g \quad (33)$$

$$\pi_t - \mu_p \pi_{t-1} = \kappa_p (m c_t + e_t^u) + \beta (E_t \pi_{t+1} - \mu_p \pi_t) \quad (34)$$

$$R_t = \rho_R R_{t-1} + (1 - \rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^R \quad (35)$$

$$b_t = \frac{1}{\beta} [(1 - \phi_b) b_{t-1} + (1 - \phi_g) e_t^g] \quad (36)$$

$$t_t = \phi_b b_{t-1} + \phi_g e_t^g \quad (37)$$

Equations (24)-(25) describe the dynamics of Tobin's  $q$ , its relationship with investments  $i_t$  and  $e_t^{rp}$  is a risk premium shock. The log-linearized law of motion of capital is in equation (26). Equation (27) is the Euler equation for  $c_t^o$ , the consumption of optimizing agents. Consumption

of the non-Ricardian agents,  $c_t^r$ , is determined by their labor income from supplying  $n_t^r$  hours at wage  $w_t$ , net of paying taxes  $t_t^r$ , as in equation (28). With flexible labor markets, the labor supply schedule for each group is in equation (29). Cost minimization implies (30) and (31), where  $mc_t$  is real marginal cost,  $e_t^z$  a total factor productivity shock and  $r_t^k$  the rental rate of capital. Output is produced as in (32). Market clearing requires that output is absorbed by aggregate consumption  $c_t$ , investment  $i_t$  and government spending  $e_t^g$ , which is random. The new Keynesian Phillips curve is in equation (34) where  $e_t^u$  is an iid markup shock,  $\mu_p$  parameterizes the degree of indexation and  $\kappa_p$  is defined as in the previous example. The central bank conducts monetary policy according to the rule (35) and  $e_t^R$  a monetary policy shock. The government budget constraint together with the fiscal rule gives rise to equation (36), where  $b_t$  denotes real bonds. The fiscal rule is in (37). In the aggregate,  $c_t = \lambda c_t^r + (1 - \lambda)c_t^o$ ,  $n_t = \lambda n_t^r + (1 - \lambda)n_t^o$ ,  $t_t = \lambda t_t^r + (1 - \lambda)t_t^o$  where  $\lambda$  is the share of non-Ricardian agents and  $t_t^j = \frac{T_t^j - T^j}{Y}$ ,  $j = o, r$ .

Before examining the question of interest, we address one preliminary issue. Does this class of models produce, with high probability, instantaneously positive consumption responses to government spending shocks when the share of non-Ricardian consumers (ROTC) is large? We draw parameters values uniformly over the intervals presented in the third column of table 7, except for  $\lambda$  which we fix at different values. The first panel of figure 6, which reports the percentage of cases in which instantaneous consumption responses to government spending shocks are negative for different  $\lambda$ , shows that the probability of finding positive consumption responses increases with the share of ROTC (in line with Galí et al. (2007)) but a large  $\lambda$  is insufficient to robustly produce the desired result. In fact, even when the majority of the consumers are not optimizers, there is a non-negligible probability that reasonable parameters configurations induce instantaneous negative consumption responses. Thus, while the condition is necessary, it is by no means sufficient.

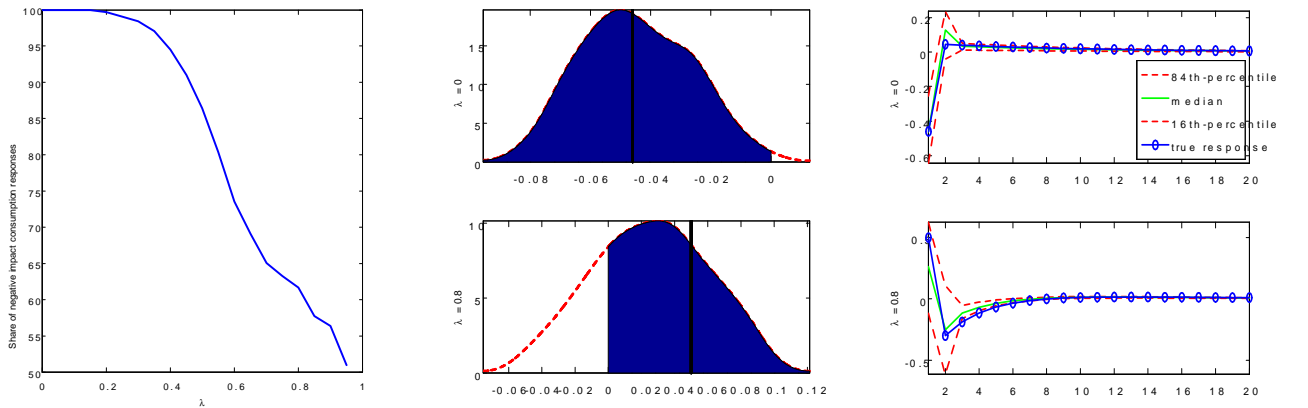


Figure 6: Consumption responses to government spending shocks, theory.

To obtain robust identification restrictions, we draw structural parameters from the intervals

presented in the third column of table 7, setting  $\beta = 0.99$  and endogenously calculating  $c_y, i_y$  using steady state conditions. The range for most of the parameters is the same as in the experiments of section 3. For the fiscal parameters, we choose arbitrary but large intervals centred around the values used Gali et al. (2007). We draw a large set of structural parameter vectors and keep only those draws producing a determinate rational expectations equilibrium - indeterminacy may occur for certain combinations of  $\lambda$  and  $\theta$ .

Parameter	Description	Support	
$\lambda$	Share of ROTC	[0.00,0.90]	0, 0.80
$\phi$	Wage elasticity to hours, ROTC	[0.00,5.00]	0.2
$\delta$	Depreciation of capital	[0.00,0.05]	0.025
$\alpha$	Capital share	[0.30,0.40]	0.33
$\eta$	Elasticity of i/K to q	[0.50,2.00]	1.0
$\theta$	Price stickiness	[0.00,0.90]	0.75
$\mu$	gross monopolistic markup	[1.10,1.30]	1.2
$\rho_r$	inertia in monetary policy	[0.00,0.90]	0.0
$\gamma_\pi$	policy response to inflation	[1.05,2.50]	1.5
$\gamma_y$	policy response to output	[0.00,0.50]	0.0
$\mu_p$	indexation in price setting	[0.00,0.80]	0.0
$\phi_b$	fiscal rule response to bonds	[0.25,0.40]	0.33
$\phi_g$	fiscal rule response to expenditure	[0.05,0.15]	0.1
$\rho_g$	AR(1) parameter gov. spending	[0.50,0.95]	0.9
$\rho_t$	AR(1) parameter productivity	[0.50,0.95]	0.9
$g_y$	steady state spending share in output	[0.15,0.20]	0.2
$\sigma_u$	standard deviation of markup shocks		0.30
$\sigma_e$	standard deviation of monetary shocks		0.025
$\sigma_z$	standard deviation of TPF shocks		0.07
$\sigma_g$	standard deviation of government shocks		0.10
$\sigma_e$	standard deviation of markup shocks		0.30
$\sigma_{rp}$	standard deviation of risk premium shocks		0.01

Table 7: Supports for the structural parameters.

#### 4.2.2 The identification restrictions

Table 8 presents the sign of the 68 percent impact response intervals of output growth, inflation, hours growth and investment growth to the five shocks. The combination of signs these intervals display is sufficient to mutually distinguish markup, technology and government spending disturbances while monetary and risk premium shocks can not be separately identified.

	Markup	Policy	Technology	Spending	Risk premium	Monetary
$\Delta y$	-	-	+	+	+	+
$\pi$	-	-	-	+	+	+
$\Delta n$	-	-	-	+	+	+
$\Delta i$	+	-	+	-	+	+

Table 8: Sign of the impact response intervals to shocks.

Before conducting our testing exercise, it is also useful to study whether our approach can distinguish situations where there are non-Ricardian consumers and where there are none, using artificial data from this class of models and the restrictions presented in table 9. In the simulation, we use the parameter values presented in the last column of table 8 (which are the same as in Gali et al. (2007)), assume the researcher observes data on output growth, inflation, hours growth, investment growth and consumption growth and that the population VAR representation of these variables is known. For illustration purposes, we consider two polar cases: no ROTC,  $\lambda = 0$ ; a large portion of ROTC  $\lambda = 0.8$ . We then ask whether the restrictions present in table 9 allow us to sign the impact consumption responses to government spending shocks with high probability and whether the dynamic responses of consumption growth in the VAR and in theory look similar.

The second panel of Figure 6 shows that in 99.6 percent of the accepted draws consumption falls on impact when  $\lambda = 0$  and in 78.2 percent of the accepted draws consumption increase on impact when  $\lambda = 0.8$  (the vertical bar in each graph denotes the true value). Furthermore, the median response path of consumption growth tracks the actual response almost perfectly in both cases (see third panel of figure 6). Hence, the method works well if the class of models has generated the data we observe and if model-based restrictions are employed for identification purposes.

#### 4.2.3 Testing the relevance of the friction and estimating $\lambda$

We estimate a 5 variable BVAR with a very loose Normal Inverted-Wishart prior using quarterly U.S. data from 1954:1 to 2007:2 obtained from the FRED database. The lag length of the VAR is set to two - this is the value selected with BIC. The BVAR includes output growth, GDP inflation, and the growth rate of hours worked in the nonfarm business sector, of private investment and of private consumption. We identify the four shocks imposing the impact restrictions appearing in table 9. We jointly draw from the posterior distribution of the BVAR parameters and orthonormal matrices until 1000 draws that satisfy all the restrictions are found.

Figure 7 presents the responses of consumption growth to government spending shocks in the data. Two interesting points can be made: when model based robust restrictions are imposed, consumption growth increases in response to a spending shock. The increase is initially large but



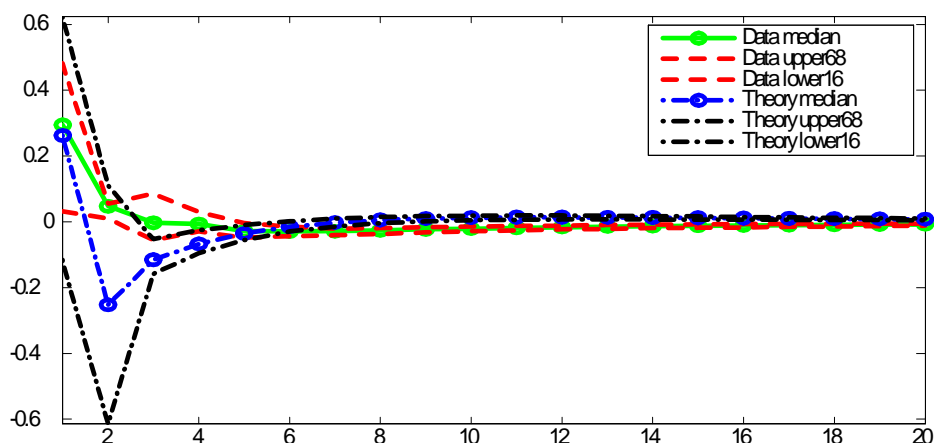


Figure 7: Estimated consumption growth responses to government spending shocks.

very short lived. Second, the distribution of the time path of the responses at different horizons in the data is similar to the distribution of responses presented in the third panel of figure 6 when  $\lambda = 0.8$  (which is superimposed in figure 7 for comparison). In fact, for the first few horizons the median of the two distributions have similar size and shape and the theory bands contain the data band. Thus, if there is interest in doing so, one could update the ranges of the intervals presented in table 8 using the information provided by the data responses and indirectly "estimate" the share of non-Ricardian consumers, something which is impossible to do with standard techniques, because  $\lambda$  can not be identified with aggregate data.

To conclude, in this class of models having a large share of ROTC is generally insufficient to produce positive consumption responses to government spending shocks. However, there are robust restrictions one can use to identify spending disturbances and to measure the sign, the magnitude and the shape of consumption responses in the data. All in all, conditional on the class of models, the share of ROTC needed to match the data dynamics is unrealistically large (see Uhlig (2009) for a similar point when fiscal multipliers are used to match the theory and the data) and this calls into serious question the use of this class for policy analyses and interpretation exercises.

## 5 Summary and conclusions

This paper presents a new methodology to examine the validity of business cycle models and to discriminate sub-models in the class. The approach employs the flexibility of SVAR techniques against model misspecification, the insights of computational experiments, and pseudo-Bayesian

predictive analysis to link the class of models to the data and exploits the computational convenience of Monte Carlo techniques to design probabilistic measures of economic discrepancy which provide effective information for model builders and applied researchers.

The starting point of the analysis 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. A subset of the robust restrictions is used to identify structural disturbances; and another subset to measure the discrepancy between the class and the data or to discriminate members of the class. In the controlled experiments we run, we found that the approach can recognize the qualitative features of DGP with high probability and can tell apart sub-models which are local to each other. It also provides a good handle of the quantitative features of the DGP if identification restrictions are abundant; and if the relative variance signal of the shock(s) one wishes to identify is sufficiently strong. The methodology is successful even when the VAR is misspecified relative to the time series model implied by the aggregate decision rules and when sample uncertainty is present.

We regard our methodology advantageous in several respects. First, it can be used even when the true DGP is not a member of the class of models one considers. Second, it does not require the probabilistic structure to be fully specified to be operative. Since misspecification is a generic feature of current business cycle models, these two characteristics crucially distinguish our approach from the existing ones. Third, our procedure de-emphasizes the quest for a good calibration and shields researchers against omitted variable biases and representation problems. Fourth, the approach can be used in a more or less limited information mode and requires limited computer time. Finally, the methodology may be turned into an interval estimation procedure for parameters that otherwise would be non-identified with standard econometric techniques.

The examples we have presented clearly indicate the potentials that the methodology has, the type of information it provides, and the interaction between theory and empirical work it produces, an interaction which is largely absent from existing methods. Recent work by Dedola and Neri (2007), Pappa (2009) among others, demonstrates that a number of interesting questions can be addressed within the framework we propose and that the answers it provides are useful for both academics and policymakers.

## 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.

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

Canova, F. 2009. Bridging cyclical DSGE models and the data, manuscript.

Canova, F. and Sala, L., 2009. Back to square one: identification issues in DSGE models. *Journal of Monetary Economics*, 56, 431-449.

Canova, F. and Gambetti, L., forthcoming. Do expectations matter? The Great Moderation revisited. *American Economic Journal: Macroeconomics*

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., 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.

Chari, V.V., Kehoe, P. and McGrattan, E., 2007. Business cycle accounting. *Econometrica*, 75, 781-836.

Chari, V.V., Kehoe, P. and McGrattan, E., 2008. Are structural VAR with long run restrictions useful for developing business cycle theory?. *Journal of Monetary Economics*, 55, 1337-1352.

Chari, V.V., Kehoe, P. and McGrattan, E., 2009. New Keynesian Models: Not yet useful for policy analyses. *American Economic Journals: Macroeconomics*, 1, 781-836.

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 Macroeconomic Annual 2006*, 1-80.

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*, 54, 512-549.

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., 2008. Forming Priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics*, 55, 1191-1208.

Del Negro, M. and Schorfheide, F., 2009. Monetary Policy Analysis with Potentially Misspecified models. *American Economic Review*, 99, 1415-1450.

Del Negro, M, Schorfheide, F., Smets, F. and Wouters, R. 2007. On the fit of New Keynesian models. *Journal of Business and Economic Statistics*, 25, 123-162.

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.

Fukac, M. and Pagan, A., forthcoming. Limited Information Estimation and Evaluation of DSGE models. *Journal of Applied Econometrics*.

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

Gali, J., and Rabanal, P. (2004). Technology shocks and Aggregate Fluctuations: How well does the RBC model fit post war US data?. *NBER Macroeconomic annual 2004*, 225-288.

Kydland, F. and Prescott, E., 1996. The Computational Experiment. *Journal of Economic Perspectives*, 10, 69-85.

Kilian, L., 1999. Small Sample confidence Intervals for Impulse Response Functions, *Review of Economics and Statistics*, 80, 218-230.

Kilian, L. and Murphy, D., 2009. Why agnostic sign restrictions are not enough: understanding the dynamics of oil markets VAR models, manuscript.

Kocherlakota, N., 2007. Model fit and model selection, *Federal Reserve Bank of St. Louis Review*, 89, 439-450.

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.

Manski, C. and Nagy, D., 1998. Bounding Disagreement about Treatment Effects: A case study of Sentencing and Recidivism. *Sociological Methodology*, 28, 99-137.

Moon, K. and Schorfheide, F., 2009. A Bayesian Look at Partially Identified Models, NBER working paper 14882.

Pappa, E., 2009. The Effects of Fiscal Shocks on Employment and Real Wages. *International Economic Review*, 50, 217-244.

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. *NBER Macroeconomic Annual 2007*, 169-226.

Rabanal, P., and Rubio-Ramirez, J., 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.

Rubio-Ramirez, J., Waggoner, D. and Zha, T. , forthcoming. Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference, *Review of Economic Studies*.

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.

Uhlig, H. 2009. Some Fiscal Calculus. University of Chicago, manuscript.

## Appendix

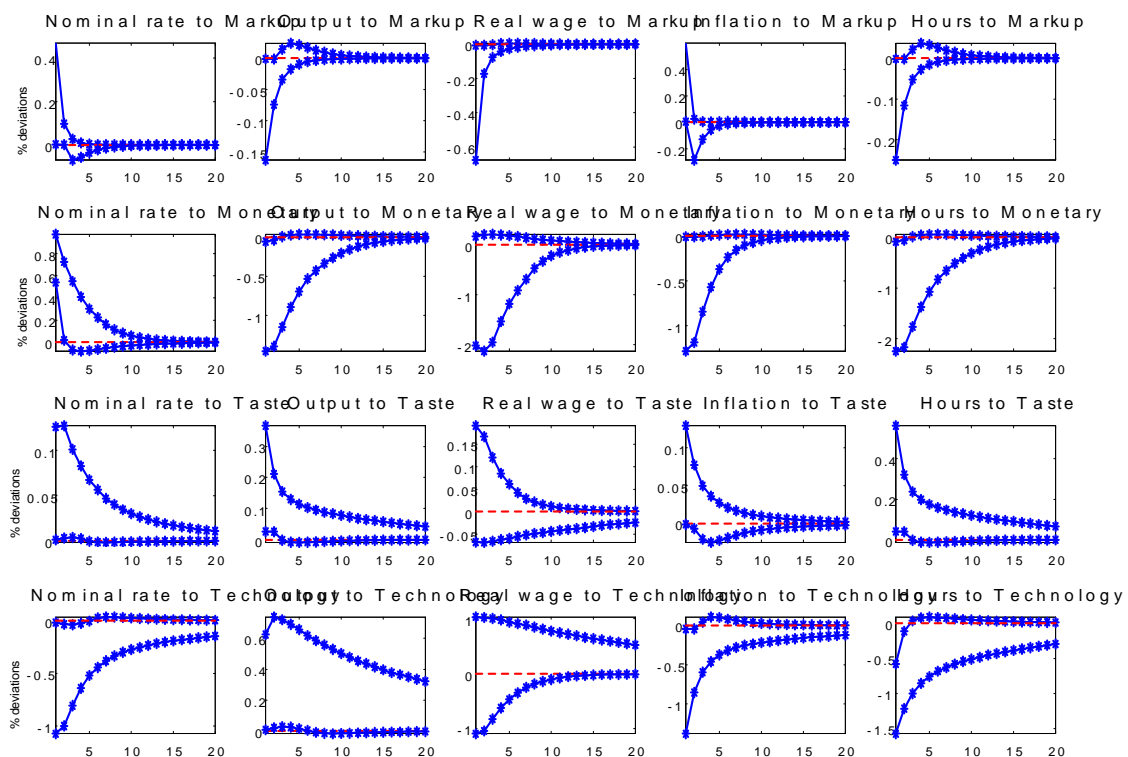


Figure A1: Pointwise 90 percent response intervals in the general model

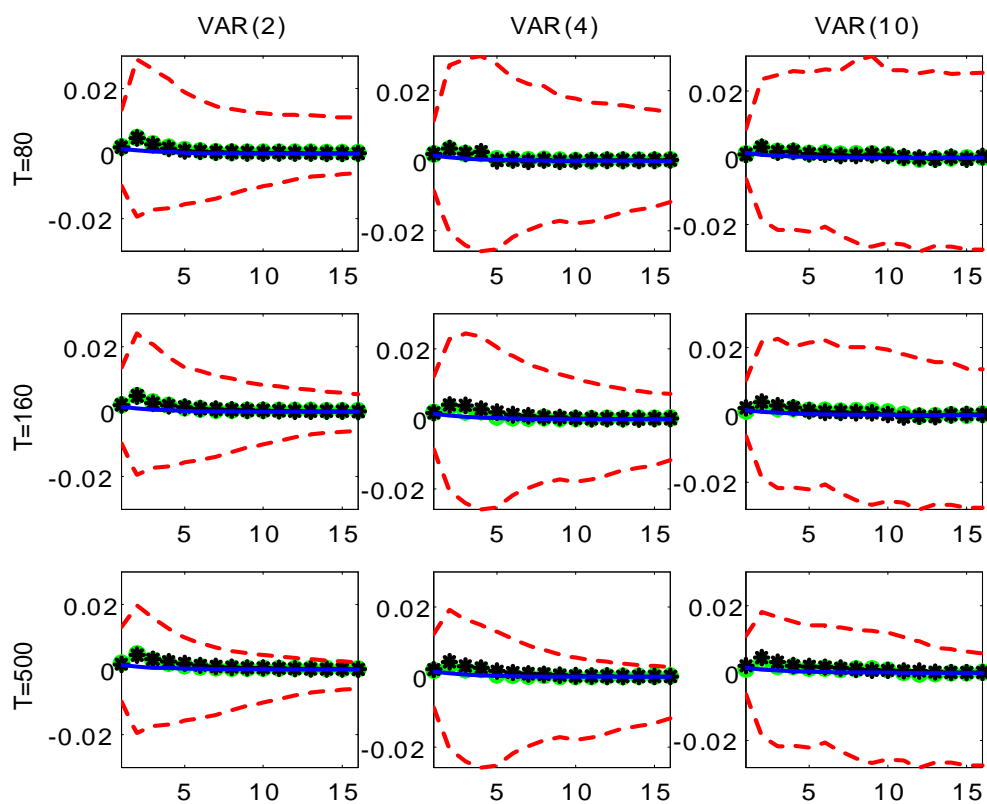


Figure A2: Pointwise 68 percent real wage response intervals to monetary shocks, all shocks identified.

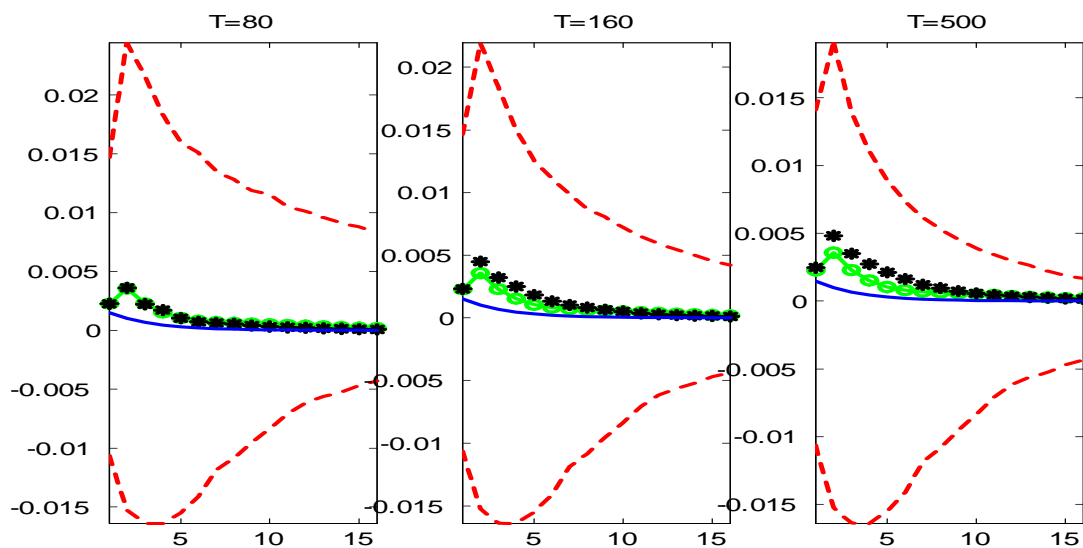


Figure A.3: Pointwise 68 percent real wage response intervals to monetary shocks, VAR chosen with BIC.



	Basic			Larger monetary shocks			Larger markup shocks		
Identified shocks	(b)	(c)	(d)	(b)	(c)	(d)	(b)	(c)	(d)
Monetary	80.9		79.4	94.8		88.6	78.3		77.4
Taste	98.3			99.1			98.0		
Supply	99.9	99.4		100	100		100	100	

Table A.1: Percentages of correct sign for the impact response of real wages in a four variable VAR. The VAR includes output, real wages, hours, inflation and the nominal rate. In b) output, inflation and nominal rate are restricted and supply, monetary and taste shocks are jointly identified, in c) and d) output, nominal rate and inflation are restricted, and either a supply shock or a monetary shock are separately identified.

	All identified			Monetary shocks identified		
	T=80	T=160	T=500	T=80	T=160	T=500
VAR(2)	62	63	64	64	64	66
VAR(4)	60	62	64	61	62	64
VAR(10)	60	61	65	60	62	65

Table A.2: Percentage of correct sign for the impact response of the real wage to monetary shocks, median value across 200 Monte Carlo replications. The DGP is a flexible price, sticky wage model and the VAR includes output, real wages, hours, inflation and the nominal rate. VAR(p) refers to the lag length of the VAR.

	Standard deviation of Monetary Shocks 10 times larger			Standard Deviation of Markup shocks 10 Times larger			Sticky price Flexible wage Model		
Horizon	T=80	T=160	T=500	T=80	T=160	T=500	T=80	T=160	T=500
0	90	90	90	50	50	49	100	100	100
1	80	86	88	64	74	77	75	85	94
2	78	80	86	64	73	81	68	76	83
3	72	76	83	62	73	80	61	69	78
4	68	72	81	59	72	83	59	63	70

Table A.3: Percentages of correctly signed real wage responses to monetary shocks; median value across 200 Monte Carlo replications. The DGP in the first two panels is a flexible price, sticky wage model and the VAR has two lags and includes output, real wages, hours, inflation and the nominal rate. The DGP in the last panel is a sticky price, flexible wage model and the VAR has two lags and includes output, real wages, hours, inflation and the nominal rate.

	2 lags			4 lags			10 lags		
Horizon	T=80	T=160	T= 500	T=80	T=160	T= 500	T=80	T=160	T= 500
0	100	100	100	100	100	100	100	100	100
1	82	89	97	78	86	95	76	87	96
2	75	78	90	63	66	85	60	71	83
3	65	69	84	53	59	70	52	57	72
4	60	61	76	59	55	63	47	54	59

Table A.4: Percentages of correctly signed real wage responses to monetary shocks; median value across 200 Monte Carlo replications. The DGP is the sticky prices, flexible wage model; the VAR includes output, inflation, nominal rate and hours. The correct representation of the DGP is a VAR(2).