barcelona gse

graduate school of economics

Measurement with Some Theory: a New Approach to Evaluate Business Cycle Models (with appendices)

Fabio Canova
Matthias Paustian

October 2010

Barcelona Economics Working Paper Series

Working Paper nº 511

Measurement with some theory: a new approach to evaluate business cycle models *

Fabio Canova[†] and Matthias Paustian[‡]

First draft: November 2007;
This draft October 13, 2010

Abstract

We propose a method to evaluate cyclical models which does not require knowledge of the DGP and the exact specification of the aggregate decision rules. We derive robust restrictions in a class of models; use some to identify structural shocks in the data and others to evaluate the class or contrast sub-models. The approach has good properties, even in small samples and when the likelihood is misspecified. We show how to sort out the relevance of a certain friction (the presence of rule-of-thumb consumers) in a standard class of models.

JEL classification: E32, C32.
Keywords: Sign restrictions, shock identification, model validation, misspecification.

---

*We thank an anonymos referee, Heather Anderson, Luca Benati, Adrian Pagan, Frank Schorfheide for comments and the participants of seminars at CREI, ECB, University of Zurich, Bank of Hungary, Riksbank, Banque of France, the conferences "How much structure in empirical models" , Barcelona, and "Monetary policy in open economies", Sydney, the 6th Dynare conference 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 not necessarily those of the Bank of England.

†ICREA-UPF, CREI, CREMeD, CEPR; Department of Economics, UPF, Ramon Trias Fargas 25-27, 08005, Barcelona (Spain), fabio.canova@upf.edu

‡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 is due to their attractive theoretical aspects, to the good empirical performance, and to the useful forecasting properties they possess, 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 and questions about their policy invariance have been raised (see e.g. Chari et al., 2009, or Chang et al., 2010); ad-hoc frictions are routinely added to match patterns found in the data, and crucial properties are derived without any reference to parameter or model uncertainty. Empirically, the problems are numerous and varied. Model misspecification is an important concern for classical estimation and generates numerical difficulties for Bayesian estimation. Identification problems make results difficult to interpret (see Canova and Sala, 2009, Iskrev, 2007, and Canova and Gambetti, 2010). The severe mismatch between theoretical and empirical concepts of business cycles (see Canova, 2009), on the other hand, renders 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. If we exclude a few notable exceptions (Schorfheide and Del Negro, 2004, and, 2009), existing work relies on likelihood ratio statistics or marginal likelihood comparisons. Both approaches focus on statistical fit rather than fundamental economic differences, are sensitive to misspecification of aspects of the models not directly tested and are computationally intensive.

This paper presents a methodology to validate classes of potentially misspecified business cycle models and to select sub-models in a class. The approach does not rely on statistical measures of fit and thus does not require estimation of often weakly identified structural parameters. Instead, 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 probabilistically evaluate the class, to discriminate among locally alternative DGPs and provide information useful to respecify theoretical structures, if needed. Dedola and Neri (2007), Pappa (2009), Peersmann and Straub (2009), Lippi and Nobili (2009) among others, have used the methodology we describe to answer interesting economic questions. What this paper provides is a formal presentation of the methodology, an assessment of its properties in simple experimental designs, and an application studying the role of rule-of-thumb consumers in generating realistic consumption responses to government expenditure shocks.

Our 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 and alternative specifications of non-essential (nuisance) aspects of the class. While magnitude restrictions depend on specification details, the sign of the impact responses is much more robust to parameter and specification uncertainty. We use a subset of theoretically robust restrictions to identify structural disturbances in the data and employ the dynamic responses of unrestricted variables to evaluate the discrepancy between the class and the data or to select a member within the class.

Our methodology has a number of advantages. First, it allows for misspecification in the structure to affect the likelihood function as long as it leaves the direction of the responses used for identification and testing unchanged. Thus, it is applicable to a richer class of problems than existing methods. Second, it can be employed to validate classes of models featuring less endogenous variables than shocks or rudimentarily specified dynamics - ad-hoc dynamics or potentially non-structural shocks need not be added for the approach to be operative. Third, by focusing shock identification and model testing on robust model-based qualitative restrictions, our methodology gives economic content to identification restrictions used in SVARs analyses and de-emphasizes the quest for good calibrations. Fourth, the procedure does not require maximization of the likelihood or the computation of the marginal likelihood, two time consuming processes, and allows researchers to make identification and testing stronger or weaker depending on the needs of the analysis.

We show that the approach can recover the sign of the impact response of unrestricted variables to the identified shocks, capture the qualitative features of the conditional dynamics, and exclude potentially relevant candidate DGPs with high probability for relevant structural designs, even when sample uncertainty exists. Moreover, it delivers reasonable conclusions even when the empirical model is misspecified relative to the DGP or the chosen class leaves important aspect of the DGP out. Finally, it can distinguish sub-models in situations where standard approaches fail.

We illustrate how the methodology can be used to gauge the frictions consistent with the observed transmission mechanism using the class of models with a portion rule-of-thumb agents, suggested by Gali et al. (2007). We demonstrate that the presence of a large number of non-optimizing consumers is insufficient to make consumption responses to government spending shocks positive. We also show how the robust restrictions the theory imposes 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 match the qualitative and quantitative features of conditional consumption dynamics in the data is unrealistically large, the validity of this class of models is seriously called into question.

The rest of the paper is organized as follows. Section 2 presents an example illustrating the robust restrictions and the testable implications a class of models delivers. Section 3 describes the testing methodology. Section 4 studies the properties of the procedure in controlled experiments. Section 5 evaluates a particular class of business cycle models. Section 6 concludes.

2 From the theory to the data

To illustrate the fundamental restrictions a theoretical structure imposes on the data, we consider the class of New-Keynesian models without capital, employed e.g. by Erceg et. al. (2000), Rabanal and Rubio Ramirez (2005) among others, which allows for habit in consumption and for price and wage rigidities.

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

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

$$\lambda_t = e_t^b - \frac{\sigma_c}{1-h}(y_t - hy_{t-1}) \tag{2}$$

$$y_t = e_t^z + (1-\alpha)n_t \tag{3}$$

$$mc_t = w_t + n_t - y_t \tag{4}$$

$$mrs_t = -\lambda_t + \sigma_l n_t \tag{5}$$

$$w_t = w_{t-1} + \pi_t^w - \pi_t \tag{6}$$

$$\pi_t^w - \mu_w\pi_{t-1} = \kappa_w\left[mrs_t - w_t\right] + \beta(E_t\pi_{t+1}^w - \mu_w\pi_t) \tag{7}$$

$$\pi_t - \mu_p \pi_{t-1} = \kappa_p \left[mc_t + e_t^{\mu} \right] + \beta (E_t \pi_{t+1} - \mu_p \pi_t) \quad (8)$$

$$R_t = \rho_R R_{t-1} + (1-\rho_R) \left[\gamma_\pi \pi_t + \gamma_y y_t \right] + e_t^R \quad (9)$$

Equation (1) is the Euler equation: λ_t is the marginal utility of consumption, R_t the nominal interest rate, and π_t price inflation. The marginal utility of consumption with external habit is defined in equation (2) and e_t^b is a preference shock. The production function is in (3); e_t^z is a productivity disturbance and n_t are hours worked. Real marginal costs mc_t are defined in (4), and w_t is the real wage. Equation (5) gives an expression for the marginal rate of substitution, mrs_t. Equation (6) links real wage to the difference between nominal wage and price inflation. The wage and price Phillips curves with Calvo pricing are in (7) and (8). μ_p (μ_w) parameterizes the degree of backward-lookingness in price (wage) setting; e_t^{μ} is a price markup shock, and π_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, where ϵ is the steady state markup. The policy rule is in (9). The four disturbances $(e_t^z, e_t^b, e_t^R, e_t^{\mu})$ are mutually uncorrelated, mean zero processes. The productivity shock e_t^z and the preference shock e_t^b have autocorrelation coefficients ρ_z and ρ_b, respectively. The monetary shock e_t^R and the markup shock e_t^{μ} are iid. The standard deviations of the innovations are $(\sigma_z, \sigma_b, \sigma_R, \sigma_\mu)$.

We wish to derive restrictions which are robust to parameter variations, independent of the specification of nuisance features, and common to the sub-models in the class to identify shocks in the data and to test the validity of the class; and restrictions which are robust to parameter variations, independent of the specification of nuisance features but different across sub-models to select members of the class.

We label M the structure represented by (1)-(9). The sub-models of interest are: a flexible price, sticky wage model ($\zeta_p = 0$) (labelled M1); a sticky price, flexible wage model ($\zeta_w = 0$) (labelled M2); a model with no indexation ($\mu_p = 0, \mu_w = 0$) (labelled M3); a model with infinitely elastic labor supply ($\sigma_l = 0$) (labelled M4). The nuisance features we focus on are the specification of habit and of nominal rigidities. In (1)-(9), habit is additive and Calvo nominal rigidities are used. As an alternative, we consider multiplicative habit (labelled N1) and quadratic adjustment costs to prices and wages (labelled N2).

To obtain robust restrictions we specify for each structural parameter a uniform distribution over an interval, chosen to be large enough to include theoretically reasonable values,

Parameter	Description	Support	DGP1	DGP2
β	Discount factor	0.99	0.99	0.99
ϵ	Elasticity in goods bundler	6	6	6
φ	Elasticity in labor bundler	6	6	6
σ_c	Risk aversion coefficient	[1.00, 5.00]	2.00	2.00
σ_l	Inverse Frish elasticity of labor supply	[0.00, 5.00]	1.74	1.74
h	Habit parameter	[0.00, 0.95]	0	0
ζ_p	Probability of keeping prices fixed	[0.00, 0.90]	0	0.75
ζ_w	Probability of keeping wages fixed	[0.00, 0.90]	0.62	0
μ_p	Indexation in price setting	[0.00, 0.80]	0	0
μ_w	Indexation in wage setting	[0.00, 0.80]	0	0
α	1 - labor share in production function	[0.30, 0.40]	0.36	0.36
ρ_r	Inertia in Taylor rule	[0.25, 0.95]	0.74	0.74
γ_y	Response to output in Taylor rule	[0.00, 0.50]	0.26	0.26
γ_π	Response to inflation in Taylor rule	[1.05, 2.50]	1.08	1.08
ρ_z	Persistence of productivity	[0.50, 0.99]	0.74	0.74
ρ_b	Persistence in taste process	[0.00, 0.99]	0.82	0.82
σ_z	Standard deviation of productivity		0.0388	0.0388
σ_μ	Standard deviation of markup		0.0316	0.0316
σ_b	Standard deviation of preferences		0.1188	0.1188
σ_r	Standard deviation of monetary		0.0033	0.0033
σ_m	Standard deviation of measurement error		0.0010	0.0010

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

existing structural estimates or values used in calibration exercises - see third column of Table 1 [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 employed in the asset pricing literature (see e.g. Bansal and Yaron (2004)), while the intervals for stickiness and indexation parameters include, roughly, the universe of possible values considered in the literature. We then 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. Figure 1 shows the range of dynamic outcomes for the nominal rate R_t, the real wage w_t, the price inflation rate π_t, output y_t, and hours n_t for model M.

The magnitude of the responses depends on the parameterization. The sign of several

[1] The discount factor β and the elasticity parameters ϵ and φ are kept fixed as they not separately identified - they enter the two Phillips curves as composites, together with the price and wage stickiness parameters.

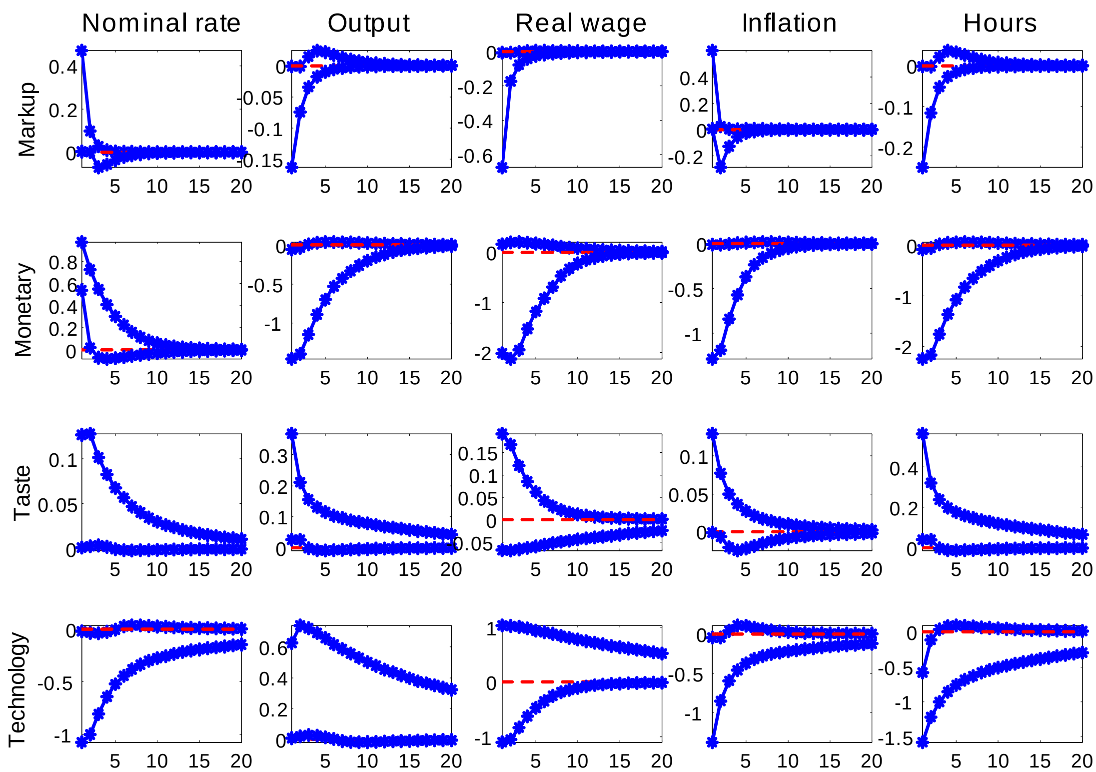

Figure 1: Pointwise 90 percent response intervals in the general model M.

dynamic responses is also fragile: the zero line is often included in the interval at medium and long horizons. The sign of impact responses is instead typically robust to the parametrization. For example, in response to markup shocks, the impact response intervals for the nominal rate and inflation are positive and those for output, real wage and hours are negative.

Are the sign of the impact response intervals independent of the specification of nuisance features? Do they hold in sub-models of interest? Table 2 reports the sign of the impact intervals in the general model, in the four submodels of interest, and in each of the two alternative specifications of nuisance features; a '+' ('-') indicates robustly positive (negative) responses; a '?' non-robust responses.

	Markup shocks							Monetary shocks						
	M	M1	M2	M3	M4	N1	N2	M	**M1**	**M2**	M3	M4	N1	N2
R_t	+	+	+	+	+	+	+	+	+	+	+	+	+	+
w_t	-	-	-	-	-	-	-	?	**+**	**-**	?	?	?	?
π_t	+	+	+	+	+	+	+	-	-	-	-	-	-	-
y_t	-	-	-	-	-	-	-	-	-	-	-	-	-	-
n_t	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	Taste shocks							Technology shocks						
	M	M1	M2	M3	M4	N1	N2	M	M1	M2	M3	M4	N1	N2
R_t	+	+	+	+	+	+	+	-	-	-	-	-	-	-
w_t	?	-	?	?	?	?	?	?	+	?	?	?	?	?
π_t	+	+	?	+	+	+	+	-	-	-	-	-	-	-
y_t	+	+	+	+	+	+	+	+	+	+	+	+	+	+
n_t	+	+	+	+	+	+	+	-	-	-	-	-	-	-

Table 2: Signs of the impact response intervals to shocks, different models. '+' indicates a robustly positive responses; '-' a robustly negative responses; '?' a response which is not robust. M is the general model, in M1 $\zeta_p = 0$; in M2 $\zeta_w = 0$; in M3 $\mu_p = 0$ and $\mu_w = 0$. In N1 habit is of multiplicative form and in N2 nominal rigidities are modelled with quadratic adjustment costs.

Many impact responses have robust signs, both across sub-models and alternative choices of nuisance features. For example, positive markup shocks increase production costs for any specification and parameterization we consider. To test the validity of this class of models one could use, e.g., the restrictions that markup shocks produce on nominal rate, inflation, output and real wages to identify these disturbances in the data and then examine whether the hours impact response interval is positive as theory predicts. Clearly, how many robust restrictions are used to identify and how many to test is question dependent. More identification restrictions avoid shocks confusion (for example, if only restrictions on output and inflation are used, markup and technology shocks are indistinguishable). More restrictions at the testing stage make the validation exercise sharper.

The impact response of the real wage to monetary disturbances is of interest since the sign of the interval differs for sub-models in the class featuring alternative nominal frictions. In sub-model M1 (flexible prices and sticky wages), workers are off their labor supply schedule and from the firm's labor demand schedule, $w_t = -\frac{\alpha}{1-\alpha} y_t$, making real wages positively comove contemporaneously with monetary shocks. In sub-model M2 (sticky prices, 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 are instantaneously negatively related to monetary shocks. Thus, to contrast sticky wages vs. sticky prices in the data, one could identify monetary shocks using the robust restrictions that the theory imposes on all variables but real wages and then examine whether real wages instantaneously fall or increase.

Distinguishing between sticky price and sticky wage models is difficult using unconditional measures of wage cyclicality because there are shocks which can instantaneously drive real wages up and down in each sub-model. Formal likelihood comparison may not be helpful either because the parameters regulating price and wage rigidities may be only weakly identified (see Del Negro and Schorfheide (2008) or Canova and Sala (2009)). The fundamental differences in the propagation mechanism we emphasize may help us to resolve the issue.

While we contrast submodels of a class, the methodology can also be employed to select classes of models featuring alternative transmission properties. In this case, one would derive robust restrictions for each class; estimate partially identified VARs using common restrictions; and select a candidate using restrictions differing in the two classes.

3 The mechanics of the evaluation approach

Our approach presumes that current business cycle models 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 also by Chari et al. (2009)). This misspecification will not necessarily vanish adding measurement errors or shocks, or tagging artificial dynamics to the model, making standard measures of fit inadequate. By focusing on fundamental features of the propagation of shocks and distinguishing alternatives using robust implications, our methodology sidesteps potential likelihood misspecification problems.

Next, we formally describe our approach and for this we need some notation. Let $F(w_t^s(\theta), \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, g, \mathcal{M}) \equiv F^s(\theta)$ be a set of continuous model-based functions, computable conditional on the structural disturbances ϵ_t, using models in the class $\mathcal{M}$, featuring the nuisance aspects g. $F^s(\theta)$ 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 θ 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 take the class $\mathcal{M}$ to be broad enough to include sub-models with interesting economic features. The nuisance features g are not of direct interest but may affect the time series properties of w_t^s. The class $\mathcal{M}$ is misspecified in the sense that even if there exists a θ_0 such that $\alpha_0 = \alpha_0(\theta_0)$ or $\alpha_1 = \alpha_1(\theta_0)$ or both, $F(w_t^s(\theta), \alpha_0(\theta_0), \alpha_1(\theta_0)|\epsilon_t, g, \mathcal{M}) \neq F(w_t, \alpha_0, \alpha_1|u_t)$.

Among all possible $F^s(\theta)$ functions, we restrict attention to the subset $\tilde{F}^s(\theta)$ which are robust to parameter variations and to the specification of nuisance features: 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 $sgn(F^s(\theta_1)) = sgn(F^s(\theta_2))$, $\forall\ \theta_1, \theta_2 \in [\theta_l, \theta_u]$, where sgn is the sign of F^s; θ_l, θ_u are the upper and lower range of economically reasonable parameter values and the above holds for all interesting specification of g. In addition, we require $\tilde{F}_1^s(\theta)$ to hold for all $\mathcal{M}_j \in \mathcal{M}$, while depending on what we test, $\tilde{F}_2^s(\theta)$ may contain functions whose sign does not depend on the sub-model (if generic fit is evaluated) or depends on $\mathcal{M}_j$ (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.

To compute $\tilde{F}^s(\theta)$ we follow Canova (1995), draw θ from some distribution, solve the model and store $F^s(\theta)$ at every draw. We then order the output, extract a confidence interval and check if it is entirely on one side of zero or compute the probability that $\tilde{F}^s(\theta)$ is on one side of the zero line. To impose $\tilde{F}_1^s(\theta)$ on the data we rotate the covariance matrix of the reduced form shocks Σ_u until $sgnF(w_{1t}^s(\theta), \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, g, \mathcal{M}) = sgnF(w_{1t}, \alpha_0, \alpha_1|u_t)$ where $A_0 A_0' = \Sigma_u$, $\alpha_0 = A_0 H$, $HH' = I$ and w_{1t} is the subset of w_t over which restrictions are imposed. An algorithm to efficiently rotate Σ_u is provided by Rubio et al. (forthcoming). There maybe many, one or no α_0 with the required characteristics. If no α_0 exists, one can impose the restrictions on another subset of w_{1t}, if available, or use another set of $\tilde{F}_1^s(\theta)$. If all interesting options are exhausted and still no α_0 is found, one can stop the evaluation process - the robust restrictions that the class of models impose have no counterpart in the data. When $k = 1, 2, \ldots, K$ α_0 values are found, we store all of them.

Model evaluation then consists in probabilistic statements concerning the features of $\tilde{F}_2(w_{2t}, \alpha_0, \alpha_1|u_t)$. For example, one can compute compute the probability that $sgn\tilde{F}_2(w_{2t}, \alpha_0, \alpha_1|u_t) - sgn\tilde{F}_2^s(w_{2t}^s, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, g, \mathcal{M}) = 0$ and the probability that $shp\tilde{F}_2(w_{2t}, \alpha_0, \alpha_1|u_t) -$

$shp\tilde{F}_2^s(w_{2t}^s, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, g, \mathcal{M}) = 0$, where shp is the dynamic shape of $\tilde{F}_2$, $w_{2t} \neq w_{1t}$ is a subset of w_t. Alternatively, one could compute the degree of overlap between the distribution of $\tilde{F}_2^s(\theta)$ and of $\tilde{F}_2(\alpha_0)$, where the distributions are obtained using the random draws of θ and of α_0 obtained in the previous steps. If only one α_0 is available, one useful summary statistics is the probability that $\tilde{F}_2^s(\theta) \leq \tilde{F}_2(\alpha_0)$ where θ are drawn from $[\theta_l, \theta_u]$. Simple graphical devices, such as plots of the 90% bands in theory and in the data, could also give a good idea of the likelihood of the restrictions.

If different sub-models have to be selected, one can construct, e.g., the probability that $sgn\tilde{F}_2(w_{2t}, \alpha_0, \alpha_1|u_t) - sgn\tilde{F}_2^s(w_{2t}^s, \alpha_0(\theta), \alpha_1(\theta)|\epsilon_t, g, \mathcal{M}_j) = 0$ for each $\mathcal{M}_j$ and select the model with the highest probability. Alternatively, one could plot confidence intervals for the sub-models of interest and take the one where the overlap with the theory is largest.

3.1 Discussion

To derive robust constraints, we focus on the sign of the impact responses period for two reasons: theory does not impose robust magnitude restrictions and dynamic responses do not always have robust sign; if models are misspecified, magnitude restrictions need not hold in the data. We employ conditional functions, such as impulse responses, since they are more informative than e.g. unconditional moments about the features of the class $\mathcal{M}$.

The identification process may involve more or less restrictions and one or more disturbances can be obtained. Hence, the methodology is flexible and can be adapted to the need of the analysis. Since standard rank and order conditions are not applicable to our case, how minimal this set of restrictions must be is generally unknown. Some indications on to proceed in practice are provided in the next section. Contrary to traditional practices, the identification restrictions we use are explicitly derived from a class of models and only constraints which are robust within the class are employed. Thus, we obtain generic conditional dynamics and refrain from conditioning on any particular member of the class or on its parameterization.

The evaluation process we employ is similar to the one employed in computational experiments where some moments are used to calibrate the structural parameters and others to check the goodness of the theory. Here a subset of the robust sign restrictions are employed to identify structural disturbances; the sign and the shape of the dynamic responses of unrestricted variables are used to check the quality of the model's approximation to the data or to select sub-models in the class. We differ in two important respects: we use qualitative rather than quantitative restrictions at both stages; our evaluation process is probabilistic.

Researchers are often concerned with the relative likelihood of sub-models in a class differing in terms of microfundations, frictions, or functional forms. While the likelihood function need not be informative about these differences, our approach can, whenever sub-models differ in the sign (shape) of certain responses. For example, it is well known that sticky and flexible price versions of the same class of model produce different signs for the instantaneous 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 can be used to discriminate the two theories. If sub-models differ in a number of implications, a weighted average of the relevant probabilities can be used to select the model with the smaller discrepancy with the data. Candidate sub-models could be nested and or non-nested: our method works in both setups.

Our approach compares favorably to existing methods, both of classical and of Bayesian inclination, for at least four reasons. First, the use of robust identification/testing restrictions shields researchers from model and parameter misspecification. All that the approach requires is that any misspecification leaves the sign of the impulse responses that are used for identification and testing unchanged. Clearly, we cannot rule out that some type of misspecification changes the sign of key impulse responses; but qualitative restrictions on the sign of conditional moments tend to hold across many forms of misspecification. Second, since the mapping between the structural parameters and the coefficients of the decision rules is not exploited in testing, lack of parameter identification is less of a problem in our framework. In any case, since the set of α_0 we derive is not necessarily a singleton, the procedure recognizes that the relationship between the $\alpha_i, i = 0, 1$ and the θs may not be unique. Third, our evaluation procedure is cheap computationally. Distributions of outcomes in theory are obtained when robust restrictions are sought; distributions of data outputs are obtained during the identification process. Since both require simple Monte Carlo exercises, the computational burden is much smaller than the one involved in classical or Bayesian Likelihood-based evaluation techniques. Finally, the statistics we construct can help to respecify the class of models, if the match with the data is unsatisfactory. For example, shape differences may suggest what type of amplification mechanism may be missing and sign differences the frictions that need to be introduced.

3.2 The relationship with the literature

Our methodology is related to early work by Canova, Finn and Pagan, (1994) and Canova (1995), and to the recent strand of literature identifying VAR disturbances using sign restrictions (see Canova and De Nicolo', 2002, or Uhlig, 2005). It is also related to Del Negro and Schorfheide (2004) and (2009), who use the data generated by a cyclical model as a prior for reduced form VARs. Two differences set our approach apart: we condition the analysis on a general class rather than on a single model; we only work with qualitative rather than quantitative restrictions. 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 (2010) 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. Our emphasis on model evaluation techniques which do not employ statistical measures of fit is also present in Kocherlakota (2007), who shows that the best fitting model is not necessarily the more accurate for policy and inferential exercises, when the available candidates are all misspecified.

4 The evaluation procedure in controlled experiments

To examine the properties of our procedure in realistic settings, we consider either the small scale class of models described in section 2 or the larger scale version used by Smets and Wouters (2003) as DGPs in our experiments. We proceed in two steps. First, we investigate the properties of our procedure in population. Later, we discuss whether sampling and specification uncertainty make a difference.

4.1 Population analysis

We start with the class of section 2 and pick the flexible price, sticky wage sub-model M1 as our DGP. The parameters used in simulating "pseudo-actual" data are the fourth column of table 1 and similar to the estimates of Rabanal and Rubio-Ramirez (2005). We endow the researcher with (1)-(9) and its solution, and let both the model dynamics and the covariance matrix of the reduced form errors Σ be known. We ask whether the responses of the real wage can be recovered with high probability employing different subsets of the robust restrictions, 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 matrices with zero mean, unitary variance; ii) apply the QR decomposition and construct impact responses as $\alpha_0 = S * Q$, where $SS' = \Sigma$; iii) keep the responses satisfying the restrictions we impose. To make results stable, we draw until 10000 candidates satisfying the restrictions are found.

4.1.1 Can we recover the true model?

In the baseline case, the empirical model includes 5 variables: the nominal rate, output, inflation, hours and the real wage. Since the economy features 4 structural shocks, we attach 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 setups where either the standard deviation of monetary shocks or the standard deviation of the markup shocks is 10 times larger, and for each we repeat the four experiments. Table 3 reports the percentage of correctly signed impact real wage responses.

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 attention. First, the number of shocks identified seems to matter. For instance, in the case of a 5 variable VAR and when a large standard deviation for markup shocks is assumed, we find that moving from identification scheme (d) which imposes restrictions only on responses to monetary shocks to identification scheme (a) which restricts responses to four structural shocks, raises the

	5 variable VAR															
	Basic				Larger monetary shocks				Larger markup shocks							
Identified 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						
	4 variable VAR															
	Basic				Larger monetary shocks				Larger markup shocks							
Identified shocks	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(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 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 the first panel and output, real wages, inflation and the nominal rate in the second panel. 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 (supply) or a monetary shock are separately identified. In the second and third panel the standard deviation of either the monetary or of the markup shocks is set 10 times larger.

fraction of correctly signed responses to monetary shocks by 8 percentage points. In other cases, the increase is smaller. In general, the benefit from identifying further shocks when the economic interest is only in one particular structural shock depends on the DGP and seem to be larger when the strength of various shocks is more heterogeneous.

Second, as in Paustian (2007), the relative strength of the shock signal matters. For instance, when we increase the standard deviation of the monetary shock tenfold, the fraction of correctly identified real wage responses to monetary shocks rises from about 75% to about 90% under identification scheme (d). Conversely, if the relative strength of the monetary shock signal is reduced by increasing the standard deviation of the markup shock tenfold, the fraction of correctly signed responses to monetary shocks falls from roughly 75% to roughly 50% again under identification scheme (d). On the other hand, 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.

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, their transmission properties 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.

The same conclusions hold when hours is dropped from the VAR. 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. When robust sign restrictions on the impact response are used for identification, omission of a state variable is less crucial for inference.

4.1.2 Can we exclude alternative models?

As table 2 shows, a sticky price, flexible wage sub-model (M2) and a flexible price, sticky wage sub-model (M1) are local to each other as far as the sign of impact responses is concerned. Our procedure can recover the sign of the real wage response to monetary shocks well when M1 is the DGP. Would the answer be different if M2 and the parameterization listed in the last column of table 1 characterizes our DGP? 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 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 shown 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 we present.

Why is our procedure successful in both capturing the DGP and in excluding local sub-models 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 theory has sharp and alternative implications for the real wage responses to monetary shocks. The answer would be different if the implications of different sub-models were more mudded. 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 configurations. Hence, 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) showed 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. Del Negro and Schorfheide (2008) report similar difficulties when Bayesian methods are used. Our semi-parametric approach, which does not require structural parameter estimation, can give sharp answers even when identification problems are present.

4.1.3 Summarizing the shape of the dynamic responses

So far we have used the sign of the impact response of a variable left unrestricted in the identification process to test the propagation mechanism of a sub-model. For many purposes this restricted 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 may be of interest. Alternatively, one may want to extend the testing to multiple horizons (if robust restrictions exist) and ask, for example, whether there exists a location measure that reasonably approximates, say, certain conditional multipliers.

Figure 2 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. It also captures the sign of the dynamics well, but it is an imperfect estimator of the magnitude of the conditional real wage dynamics, at least as far as the responses to monetary shocks are concerned. 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).

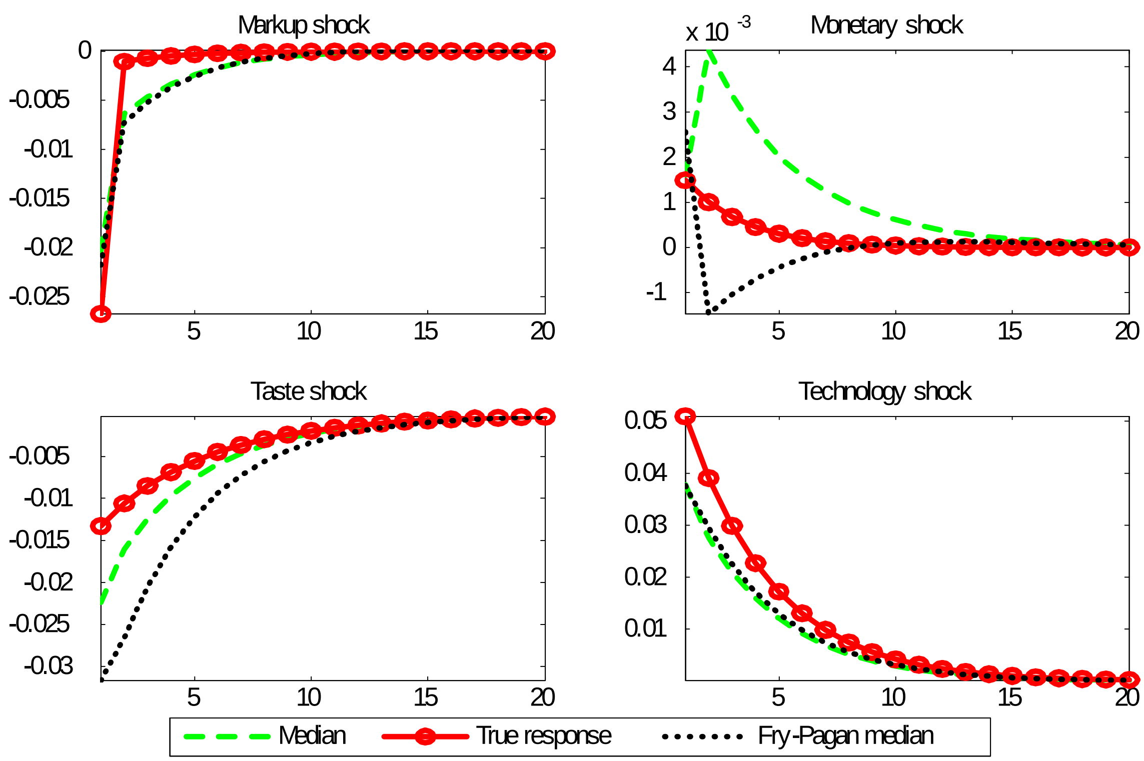

Figure 2: Real wage responses to shocks.

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, identified shocks may be correlated. 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, ranges from 0.59 to 0.89 in absolute value. Therefore, Fry and Pagan's concern seems legitimate. However, as figure 1 shows, this alternative 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. In addition, the correlation between true and estimated disturbances obtained this way is generally low and for monetary shocks it is surprisingly negative. Thus, if the magnitude dynamic responses to monetary shocks is of interest, it is unclear which measure dominates; if the sign of the dynamics is crucial, having uncorrelated shocks may be worse.

We have conducted numerous exercises to check whether the performance of the median is affected by the experimental design. We find that (i) identifying more shocks or increasing the strength of the variance signal improves the dynamic performance of the median; (ii) the dimensionality of the VAR has no influence on the dynamic properties of the median; and (iii) using model M1 or M2 as the DGP makes no difference for the conclusions we reach.

4.2 Does sampling uncertainty matter?

The ideal conditions considered so far are useful to understand the properties of the procedure but unlikely to hold in practice. What happens if the autoregressive parameters and the covariance matrix of the shocks are estimated prior to the identification exercise?

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 adds misspecification - the decision rules imply that a VAR(∞) should be used. We also 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 summarize the outcomes of our experiments in table 4 by reporting the probability that the impact response of the real wage to monetary shocks has the correct sign. Here the DGP is 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 and shocks are identified imposing sign restrictions on the impact responses of the nominal rate, output, inflation and hours. Additional statistics for this and other experiments we run are in the accompanying materials (appendix A).

	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
BIC	62	63	65	64	65	66

Table 4: 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. p is to the lag length of the VAR. The row labeller "BIC" reports probabilities computed when BIC is used to select the lag length of the VAR.

Three features of table 4 stand out. First, sample uncertainty is small relative to identification uncertainty: the probabilities we report increase with the sample size for each lag length, but the differences between T=80 and T=500 are small. Second, changing the lag length of the VAR has little consequences on the outcomes. With a larger number of lags, the probability generally falls, but the difference are remarkably small. Since, these patterns are also present when the lag length of the VAR is selected with BIC, none of the problems highlighted by Chari, et al. (2008) appear to be present here. Third, the number of shocks we identify has minor consequences on the quality of the outcomes.

All other conclusions obtained in population hold also here. For example, as shown in the accompanying materials (appendix A), the number of variables included in the VAR has little effect on the conclusions, and changing the variability of shocks produces the same results found in population. We can also still recognize the DGP and exclude local sub-models with high probability looking at the impact response of the real wage to monetary shocks.

Finally, the performance of the median, as summary measure for the true responses, is broadly unaffected. In sum, sample uncertainty is small relative to identification uncertainty (see Kilian and Murphy, 2009, for related evidence); and lag specification uncertainty has minor consequences on the performance of our approach.

4.3 Using the wrong model for inference

Although we have stated that misspecification of the likelihood function is less of a problem for our approach, it is interesting to examine what would happen if our procedure were applied to a class of models which leaves out important aspects of the true DGP. For that purpose, we simulate data from a version of Smets and Wouters (SW) (2003) class of models, and use this dataset to test the validity of the restrictions imposed by the class of models of section 2. The log-linearized optimality conditions, the parameter intervals used to derive robust restrictions and the parameters of the DGP are in the accompanying materials (appendix B).

	TFP	Monetary	Taste	Investment	Markup	Labor supply	Goverment
y_t	+	+	+	?	+	+	+
π_t	-	+	+	-	-	-	?
R_t	-	-	+	?	-	-	+
w_t	?	?	?	?	+	-	?
n_t	-	+	+	?	+	+	+
LP-W gap_t	+	?	-	+	-	-	-

Table 5: Signs of the 90 percent impact response intervals to shocks, SW class.

To begin with, we show what robust restrictions the SW class imposes on output, inflation, the nominal rate, real wages and hours for each of the seven disturbances. Table 5 reports the signs of the 90 percent impact response intervals. Interestingly, the sign of the intervals in responses to TFP, monetary, taste and markup disturbances are the same as in table 2 and are robust across sub-models. Thus, inference would not be necessarily distorted if a class models which leaves out variables and frictions present in the GDP is used to derive robust restrictions.

However, table 5 also shows that these restrictions alone would not be sufficient to uniquely obtain the four disturbances. In fact, in a four variable VAR, identified shocks may capture, in principle, any of the seven true structural shocks. For example, in our case, taste shocks could capture, in part, government expenditure shocks, while markup and technology shocks may reflect investment specific shocks. To check the extent of the problem, we have computed what is the proportion of correctly signed real wage responses to shocks in population. It turns out that some contamination is present, but it is generally small. For example, when markup, monetary, taste and technology shocks are identified using 16 impact restrictions the probabilities of correctly signing the impact real wage response are 98.1, 98.7, 90.7 and 98.8, respectively. When only three shocks are identified using 12 impact restrictions, the probabilities are 98.6 for supply shocks, 99.5 for monetary shocks and 91.0 for taste shocks.

How can one limit shock confusion? Shrewdly choosing the variables of the VAR helps. As the last row of table 5 shows, if the labor productivity-real wage gap is added and the nominal rate is dropped from the list of observables, the seven shocks produce mutually exclusive patterns of signs on the contemporaneous responses of the five variables of interest. Thus, shock confusion will be unlikely even if the smaller class of models is used for inference.

4.4 Testing multiple restrictions

With the SW DGP we can also illustrate how the use of multiple restrictions - some of which may not be directly of interest - can strengthen testing in relevant practical situations. For the class we consider, the instantaneous response of hours is robustly negative to TFP shocks, if some price rigidities are present, and robustly positive to labor supply, investment and markup shocks, regardless of the extent of price rigidities. The first implication is typically evaluated in the empirical literature, but hardly anyone seems to care about the other implications of the theory. However, when price rigidities are not strong, jointly testing the four restrictions may give sharper answers, even if the latters are not of interest. To show this, we have simulated data from the SW class using the same parameters as before except that we set $\zeta_p = 0.3$ and $\mu_p = 0$ and computed the probability that the impact response of hours is negative in response to TFP shocks and the probability that the impact response of hours is negative in response to TFP shocks and positive in response to investment, labor supply and markup shocks.

The former probability is 39 percent indicating that, when price stickiness is low, it is difficult to distinguish presence or absence of price rigidities. This probability falls to

17 percent when the four restrictions are jointly tested - the difference is due to rotations matrices that imply positive hours responses to TFP shocks but negative hours responses to any of the other three shocks. Thus, when the data does not speak very loud about the question of interest, testing a larger set of restrictions can sharpen inference.

4.5 Advice to the users

Our procedure has good properties in all the experiments we consider. 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 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. Similarly, sharper answers to the questions can be obtained if a number of robust restrictions, some which are of interest, some which are not, are jointly tested.

Our experiments also showed that credible intervals tend to be large - this expected given that the methodology delivers partially identified empirical models (see Moon and Schorfheide (2009). Nevertheless, the probabilistic summary statistics we employ are informative about the features of the DGP, even when asymptotically-based standard normal tests are not. If one insists on using the latter, a sufficient number of restrictions and smaller confidence intervals (say, 68 percent or interquartile ranges) need to be employed at the inferential stage.

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 loose but informative prior was sufficient to reduce them. Other approaches, such as 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. While pathological examples can always be constructed (see Paustian (2007) or Fry and Pagan (2007)), and the strength of the shock signal is a-priori unknown, relative variance differences become a serious problem only in 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.

If a small scale class of models is used in the analysis, the choice of variables to be included in the VAR should be guided not only by economic but also by identification considerations. If the selected variables are such that the shocks produce mutually exclusive pattern of robust signs in theory, it is unlikely that the identified shocks mix true shocks of different type, making shock aggregation issues (see e.g. Faust and Leeper, 1997) less important.

Along the same lines, it is often the case that in theory 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, when a 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. As shown in the accompanying materials (appendix C), failure to impose the uniqueness condition in identification, may lead researchers astray. Thus, unless all shocks are identified, we recommend users to always impose it.

Finally, as section 4.3 has shown, misspecification of the likelihood function does not necessarily imply wrong inference. In addition, we do not need that the class of models used to derive the restrictions has the same number of shocks as the empirical VAR. All that is required is that any shock omitted from the structural model, but present in the data, is not isomorphic to the shocks of interest. Thus, stochastic singularity is not a problem and there is no need to add ad-hoc shocks to the structural model. All in all, starting from a good fitting (large scale) class is not a precondition for the methodology to be applied.

5 An example

It is well known that standard business cycle models find it difficult to reproduce 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 this literature are not explicitly derived from any theoretical specification that it then used to interpret the results. Gali et al. (2007) have taken a standard New Keynesian class of models and showed that adding one particular friction (a portion of non-Ricardian consumers) can make the theory consistent with the existing structural VAR evidence. This section investigates three separate questions. First, does the Gali et al. class of models produce consumption responses to spending shocks which are positive with high probability? Second, how do consumption responses in the data look like if the robust sign restrictions the theory imposes are used to identify government spending shocks? Third, what is the likelihood that this class of models has generated the data?

5.1 The class of models

The log-linearized optimality conditions for the class of models ware

$$q_t = \beta E_t q_{t+1} + [1-\beta(1-\delta)]\, E_t r^k_{t+1} - (R_t - E_t\pi_{t+1}) \quad (10)$$

$$i_t - k_{t-1} = \eta q_t \quad (11)$$

$$k_t = (1-\delta)k_{t-1} + \delta i_t \quad (12)$$

$$c^o_t = c^o_{t+1} - (R_t - E_t\pi_{t+1}) \quad (13)$$

$$c^r_t = \frac{1-\alpha}{\mu c_y}(w_t + n^r_t) - \frac{1}{c_y} t^r_t \quad (14)$$

$$w_t = c^j_t + \sigma_l n^j_t \qquad j = o, r \quad (15)$$

$$r_t = mc_t + e^z_t + (1-\alpha)(n_t - k_{t-1}) \quad (16)$$

$$w_t = mc_t + e^z_t - \alpha(n_t - k_{t-1}) \quad (17)$$

$$y_t = e^z_t + (1-\alpha)n_t + \alpha k_{t-1} \quad (18)$$

$$y_t = c_y c_t + i_y i_t + g_y e^g_t \quad (19)$$

$$\pi_t - \mu_p \pi_{t-1} = \kappa_p(mc_t + e^u_t) + \beta(E_t\pi_{t+1} - \mu_p\pi_t) \quad (20)$$

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e^R_t \quad (21)$$

$$b_t = \frac{1}{\beta}[(1-\phi_b)b_{t-1} + (1-\phi_g)e^g_t] \quad (22)$$

$$t_t = \phi_b b_{t-1} + \phi_g e^g_t \quad (23)$$

Equations (10)-(11) describe the dynamics of Tobin's q, its relationship with investments i_t. The log-linearized law of motion of capital is in equation (12). Equation (13) is the

Euler equation for the consumption of optimizing agents, c_t^o. Consumption of the non-Ricardian agents, c_t^r, depends on their labor income obtained from supplying n_t^r hours at wage w_t, net of paying taxes t_t^r, where α is the share of labor in production, as in equation (14). With flexible labor markets, the labor supply schedule for each group is in equation (15). Cost minimization implies (16) and (17), where mc_t is real marginal cost, e_t^z a total factor productivity shock and r_t the rental rate of capital. Output is produced as in (18). 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 (20) where e_t^u is an iid markup shock, μ_p parameterizes the degree of indexation and $\kappa_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}$ and ζ_p is the Calvo probability of non-changing prices. The central bank conducts monetary policy according to the rule (21) and e_t^R a monetary policy shock. The government budget constraint together with the fiscal rule gives equation (22), where b_t are real bonds. The fiscal rule is in (23). 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 λ is the share of non-Ricardian agents and $t_t^j = \frac{T_t^j - T^j}{Y}$, $j = o, r$.

5.2 Evaluating the friction in theory

The literature often presumes that this class of models produces instantaneously positive consumption responses to government spending shocks when the share of non-Ricardian consumers (ROTC) is sufficiently large. Here, we want to know whether this constitutes a robust implication of the theory or not. To do this, we draw parameters values uniformly over the intervals presented in the third column of table 6, except for λ which we fix at different values. The first panel of figure 6, which reports the percentage of draws in which instantaneous consumption responses to government spending shocks are negative for different λ, shows that the unconditional probability of finding positive consumption responses increases with the share of ROTC but a large λ 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. To make consumption responses positive with high probability, we need something else. The remaining lines in the first panel of figure 6 show that if a large share of ROTC is combined with large price stickiness, the required result obtains. Thus, while a large value of λ is necessary, it is by no means sufficient. It is only when λ exceeds 0.7 and ζ_p exceeds 0.8 that we can confidently conclude (say, with at least 68 percent probability) that this class has the required feature.

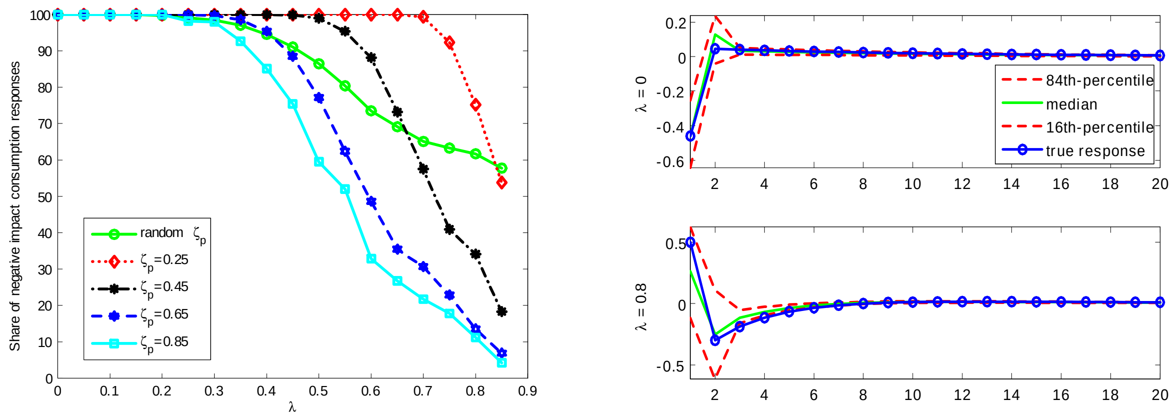

Figure 6: Consumption responses to government spending shocks, theory.

5.3 Deriving robust theoretical implications

To obtain robust identification restrictions, we draw structural parameters from the intervals presented in the third column of table 6, setting $\beta = 0.99$, endogenously calculating c_y, i_y using steady state conditions, and keeping only those draws producing a determinate rational expectations equilibrium - indeterminacy may occur for certain combinations of λ and ζ_p. The range for most of the parameters is the same as in the experiments of section 4. For the fiscal parameters, we choose large intervals centered around the values used in the literature.

Table 7 presents the sign of the 68 percent impact response intervals of output growth, inflation, hours growth, investment growth to the four shocks. The combination of signs these intervals display is sufficient to mutually distinguish all the disturbances. This would not be the case, for example, if the nominal interest rate is used in place of investment growth (markup and expenditure shocks will have similar sign implications), as it is typically the case in empirical VARs present in the literature.

Prior to the testing exercise, it is useful to check in a controlled experimental design whether our approach can distinguish situations with and without non-Ricardian consumers using the restrictions of table 7. In the simulation, we use the parameter values presented in the last column of table 6 (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.

Parameter	Description	Support	DGP
λ	Share of ROTC	[0.00,0.90]	0, 0.80
σ_l	Wage elasticity to hours	[0.00,5.00]	0.2
δ	Depreciation of capital	[0.00,0.05]	0.025
α	Capital share	[0.30,0.40]	0.33
η	Elastictiy of i/K to q	[0.50,2.00]	1.0
ζ_p	Price stickiness	[0.00,0.90]	0.75
μ	Gross monopolistic markup	[1.10,1.30]	1.2
ρ_r	Inertia in monetary policy	[0.00,0.90]	0.0
γ_π	policy response to inflation	[1.05,2.50]	1.5
γ_y	Policy response to output	[0.00,0.50]	0.0
μ_p	Indexation in price setting	[0.00,0.80]	0.0
ϕ_b	Fiscal rule response to bonds	[0.25,0.40]	0.33
ϕ_g	Fiscal rule response to expenditure	[0.05,0.15]	0.1
ρ_g	AR(1) parameter gov. spending	[0.50,0.95]	0.9
ρ_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
σ_u	Standard deviation of markup shocks		0.30
σ_R	Standard deviation of monetary shocks		0.025
σ_z	Standard deviation of TPF shocks		0.07
σ_g	Standard deviation of goverment shocks		0.10

Table 6: Supports for the structural parameters.

For illustration, we consider two polar cases: no ROTC, $\lambda = 0$; a large portion of ROTC $\lambda = 0.8$. In both cases we select $\zeta_p = 0.75$ to make the practical distinction between the two setups empirically relevant. We then ask whether the restrictions present in table 7 allow us to sign the impact consumption growth response to government spending shocks with high probability and whether the dynamic responses of consumption growth in the VAR and in theory look similar. It turns out 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 $\lambda = 0.8$. Furthermore, the median response path of consumption growth tracks the true response almost perfectly in both cases (see second panel of figure 6). Hence, the method can detect both the sign of the impact consumption responses and the shape of its dynamic responses to spending shocks, if the class of models has generated the data we

	Markup	Policy	Technology	Spending
Δy	-	-	+	+
π	-	-	-	+
Δn	-	-	-	+
Δi	+	-	+	-
R	-	+	-	+

Table 7: Signs of the impact response intervals to shocks.

observe and if model-based restrictions are employed to identify spending shocks.

5.4 Testing the relevance of the friction

We estimate a 5 variable BVAR with a 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 as selected by 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 16 impact restrictions appearing in the top of table 7. We jointly draw from the posterior of the BVAR parameters and orthonormal matrices until 1000 draws satisfying the restrictions are found.

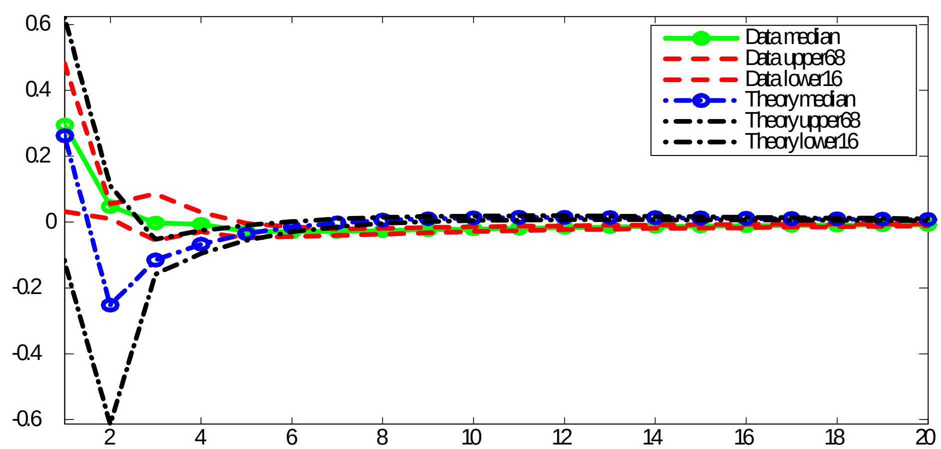

Figure 3: Consumption growth responses to goverment spending shocks.

Figure 3 presents the responses of consumption growth to government spending shocks in the data. When model based robust restrictions are imposed, consumption growth instantaneously increases. The point estimate is 0.25 and it is statistically significant but there is considerably uncertainty concerning the magnitude of the instantaneous consumption multiplier to spending shock (it could be anywhere between 0.06 and 0.6). Moreover, this increase is very short lived and after one quarter the 68 percent band includes zero. Thus, when theory-based sign restrictions are used, the instantaneous consumption response to spending shocks are comparable to those found in the micro literature for tax shocks (see e.g. Broda and Parker (2008)) and quite short lived.

Is the class of models a good candidate to explain the consumption responses we see in the data? To answer this question, we superimpose in figure 3 the consumption responses obtained from the class, conditioning on $\lambda = 0.8$ and $\zeta_p = 0.75$. Clearly, the profile of the distribution of the responses in theory and in the data is similar. Instantaneously, the median responses are very close and at short horizons the median of the two distributions have similar size and shape and the theory bands contain the data band. Thus, to match both the sign and the shape of the consumption responses we see in the data, we need considerable price stickiness and an unrealistically large share of ROTC. Since empirical evidence suggests, at best, moderate micro price stickiness and it is unlikely that about 80 percent of the US population behaves as ROTC do in these models, these results call into serious question the use of this class for inference and policy analyses [2].

6 Summary and conclusions

This paper presents a new methodology to examine the validity of business cycle models and to discriminate sub-models in a class. The approach employs the flexibility of SVAR techniques against model misspecification, the insights of computational experiments, and pseudo-Bayesian predictive analysis to link models to the data. We do not use standard measures of fit to evaluate the discrepancy: instead we design probabilistic measures which are robust to misspecifications of the likelihood function and effective in providing information

[2] As noted by Gali et. al., a model with imperfectly competitive labor markets may help to make the share of rule of thumb consumers required to generate a rise in consumption to spending shocks more realistic. However, absent data on hours worked and consumption for the two types of consumers, it is difficult to directly test an imperfectly competitive labor market against the basic specification.

useful to respecify the class.

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 and for different specifications of nuisance features. A subset of the robust restrictions is used to identify structural disturbances; 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, 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, 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 as long as the robust sign restrictions we consider are not affected by the misspecification. Second, it does not require the probabilistic structure to be fully specified to be operative. 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 adapted to the needs of the user and requires limited computer time.

Apart from the example we have presented, recent work by Dedola and Neri (2007), Pappa (2009) Peersmann and Straub (2009) Lippi and Nobili (2010) among others, indicate the potentials that the methodology possesses, the type of information it provides, and the interaction between theory and empirical work it produces. One interesting extension we are currently pursuing is transforming our evaluation approach into an estimation procedure, where the initial ranges for parameters values are updated using information similar to the one presented in section 5. This approach, which provide an indirect way to obtaining parameter intervals, could overcome many of the problems that likelihood based estimation approaches face when severe identification problems are present.

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.

Broda, C. and Parker, J., 2008. The impact of the 2008 Tax Rebates on consumer spending: preliminary evidence, mimeo.

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 De Nicoló, G., 2002. Money Matters for Business Cycle Fluctuations in the G7. *Journal of Monetary Economics*, 49, 1131-1159.

Canova, F. and Gambetti, L., 2010. Do expectations matter? The Great Moderation revisited. *American Economic Journal: Macroeconomics*, 2(3), 185-205.

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

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.

Chang,Y., Kim, S., and Schorfheide, F., 2010. Financial Frictions, Aggregation and Lucas critique, NBER working paper .

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.

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.

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

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

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., 2010. Limited Information Estimation and Evaluation of DSGE models. *Journal of Applied Econometrics*, 25, 55-70.

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‘

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

Kilian, L. and Murphy, D., 2009. Why agnostic sign restrictions are not enough: understanding the dynamics of oil markets VAR models, forthcoming, *Journal of the European Economic Association*.

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.

Lippi, F., and Nobili, A., 2010. Oil and the macroeconomy: A structural VAR analysis with sign restrictions, Bank of Italy Temi di Discussione 704.

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.

Peersmann, G. and R. Straub, 2009. Technology shocks and robust sign restrictions in a Euro area SVAR. *International Economic Review*, 50(3), 727-750.

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. , 2010. Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference, *Review of Economic Studies*, 77, 665-696.

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.

Accompanying materials

Appendix A

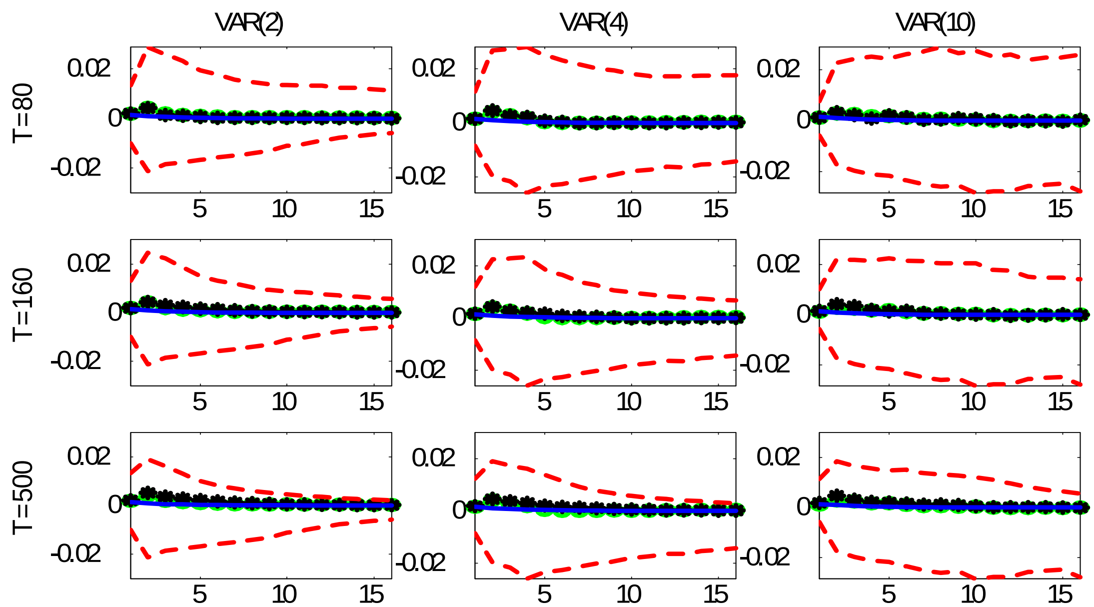

Figure A.1: Pointwise 68 percent real wage responses intervals to monetary shocks, only monetary shocks identified.

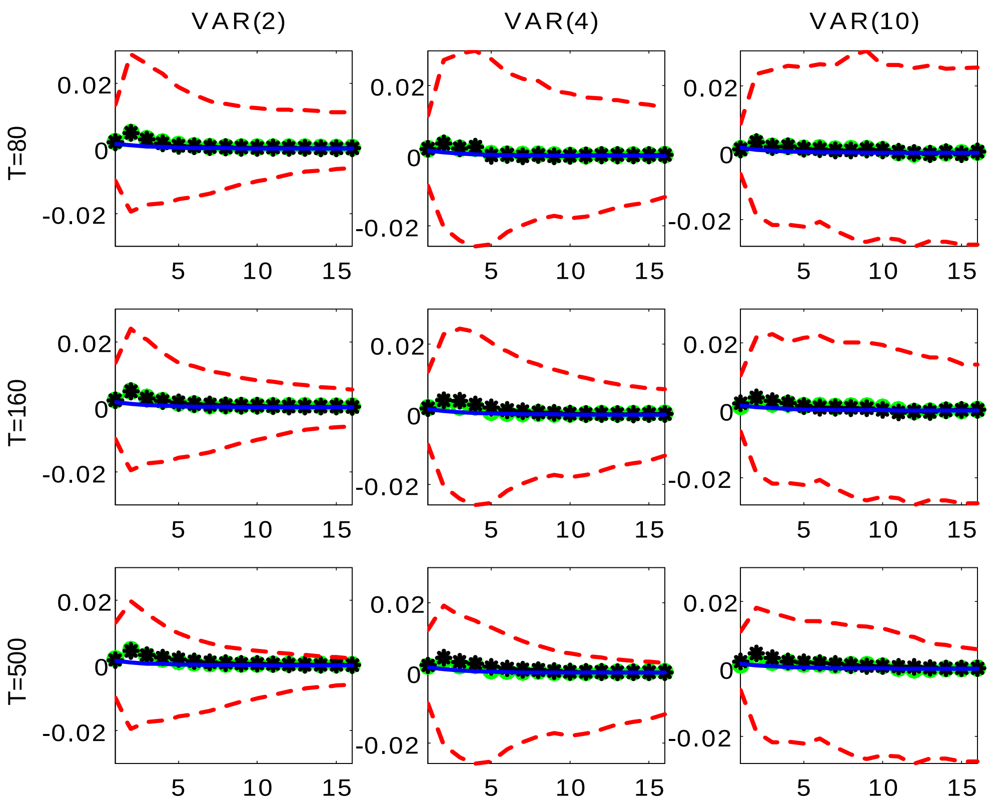

Figure A2: Pointwise 68 percent real wage response intervals to monetary shocks, all shocks identified.

825

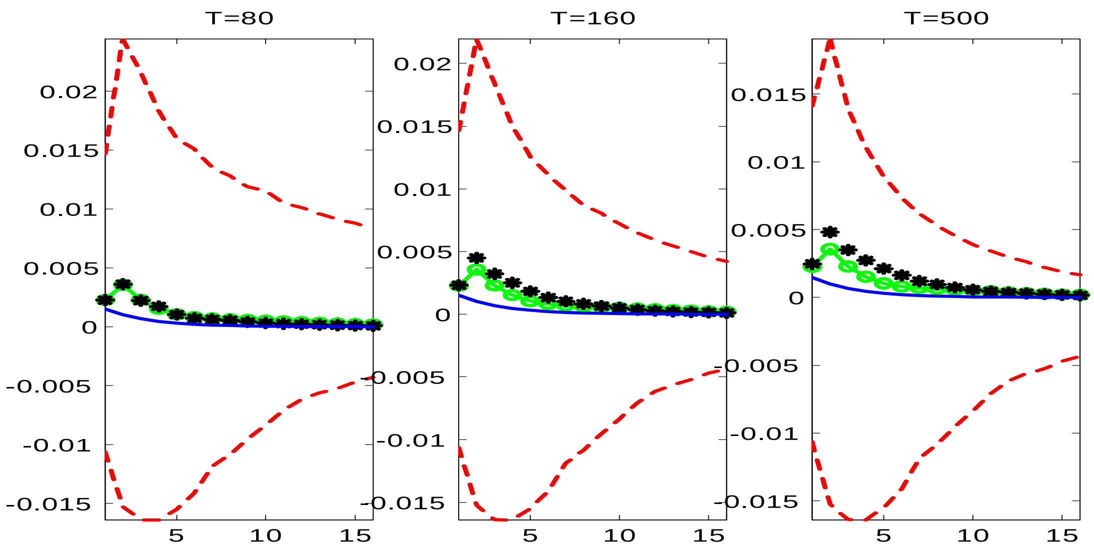

Figure A.3: Pointwise 68 percent real wage response intervals to monetary shocks, VAR chosen with BIC.

826

	Flexible price sticky wage model						Sticky price flexible price model		
	Standard deviation of monetary shocks 10 times larger			Standard deviation of markup shocks 10 times larger			Basic		
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.1: Percentages of correctly signed real wage responses to monetary shocks; median value across 200 Monte Carlo replications. In all panels 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.2: 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).

Appendix B

The Smets and Wouter's (2003) class of models 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). The class has three blocks and its log-linearized representation (around the steady state) is as follows. The aggregate demand block is:

$$y_t = c_y c_t + i_y i_t + g_y e_t^g \tag{24}$$

$$c_t = \frac{h}{1+h} c_{t-1} + \frac{1}{1+h} E_t c_{t+1} - \frac{1-h}{(1+h)\sigma_c}(R_t - E_t \pi_{t+1}) + \frac{1-h}{(1+h)\sigma_c}(e_t^b - E_t e_{t+1}^b) \tag{25}$$

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} E_t i_{t+1} + \frac{\phi}{1+\beta} q_t - \frac{\beta E_t e_{t+1}^I - e_t^I}{1+\beta} \tag{26}$$

$$q_t = \beta(1-\delta) E_t q_{t+1} - (R_t - E_t \pi_{t+1}) + (1 - \beta(1-\delta)) E_t r_{t+1} \tag{27}$$

Equation (24) 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 (25) is a dynamic IS curve: e_t^b is a preference shock, σ_c the coefficient of relative risk aversion and h the coefficient of external habit formation. The dynamics of investment are in equation (26); ϕ 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 β the discount factor. Equation (27) 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:

$$y_t = \omega(\alpha k_{t-1} + \alpha \psi r_t + (1-\alpha) n_t + e_t^z) \tag{28}$$

$$k_t = (1-\delta) k_{t-1} + \delta i_t \tag{29}$$

$$\pi_t = \frac{\beta}{1+\beta\mu_p} E_t \pi_{t+1} + \frac{\mu_p}{1+\beta\mu_p} \pi_{t-1} + \kappa_p mc_t \tag{30}$$

$$w_t = \frac{\beta}{1+\beta} E_t w_{t+1} + \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta} E_t \pi_{t+1} - \frac{1+\beta\mu_w}{1+\beta} \pi_t + \frac{\mu_w}{1+\beta} \pi_{t-1} - \kappa_w \mu_t^W \tag{31}$$

$$n_t = -w_t + (1+\psi) r_t^k + k_{t-1} \tag{32}$$

Equation (28) is the aggregate production function. In equilibrium ψr_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 α is the capital share. The law of motion of capital accumulation is in equation (29). Equation (30) links inflation to marginal costs, $mc_t = \alpha r_t^k + (1-\alpha)w_t - e_t^z + e_t^{\mu_p}$ 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 ζ_p, the probability that firms face for not being able to change prices in the Calvo setting. The parameter μ_p determines the degree of price indexation. Equation (31) links the real wage to expected and past wages, to inflation and to the marginal rate of substitution between consumption and leisure, $\mu_t^W = w_t - \sigma_l n_t - \frac{\sigma_c}{1-h}(c_t - hc_{t-1}) - e_t^{\mu_w}$, where σ_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 (32) follows from the equalization of marginal costs. The monetary rule is

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^R \tag{33}$$

where ε_t^R is a monetary policy shock.

Equations (24) to (33) define a system of 10 equations in ten unknowns, $(\pi_t, y_t, c_t, i_t, q_t, l_t, w_t, k_t, r_t, R_t)$. 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:

$$\log(S_t) = (I - \boldsymbol{\varrho})\log(\overline{S}) + \boldsymbol{\varrho}\log(S_{t-1}) + V_t \tag{34}$$

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

In table B.1 we present the intervals used to compute robust restrictions presented in table 5 together with the parameters for the DGP used in section 4.4.

Parameter	Description	Support	DGP
σ_c	Risk aversion coefficient	[1,6]	2
σ_l	Inverse Frish labor supply elasticity	[0.5,4.0]	1.9
h	Consumption habit	[0.1,0.8]	0.7
ω	Fixed cost	[1.0,1.80]	1.2
ϕ	Adjustment cost parameter	[0.0001,0.02]	0.018
δ	Capital depreciation rate	[0.015,0.03]	0.025
α	Capital share	[0.15,0.35]	0.3
ψ	Capacity utilization elasticity	[0.1,0.6]	0.5
ζ_p	Degree of price stickiness	[0.4,0.9]	0.7
μ_p	Price indexation	[0.2,0.8]	0.2
ζ_w	Degree of wage stickiness	[0.4,0.9]	0.8
μ_w	Wage indexation	[0.2,0.8]	0.5
ε^w	Steady state markup in labor market	[0.1,1.8]	1.0
g_y	Share of government consumption	[0.10,0.25]	0.2
ρ_R	Lagged interest rate coefficient	[0.2,0.95]	0.74
γ_π	Inflation coefficient on interest rate rule	[1.1,3.0]	1.18
γ_y	Output coefficient on interest rate rule	[0.0,1.0]	0.0
$\boldsymbol{\varrho}_i$	Persistence of shocks $i = z, b, I, \mu_p, \mu_w$	[0,0.9]	0.8
β	Discount factor	0.99	0.99
π^s	Steady state inflation	1.016	1.016
s_g	Standard deviation expenditure shock		0.1
s_b	Standard deviation preference shock		0.066
s_z	Standard deviation technology shock		0.0064
s_i	Standard deviation investment shock		0.557
s_p	Standard deviation price markup shock		0.221
s_w	Standard deviation wage markup shock		0.135
s_m	Standard deviation monetary shock		0.0026

Table B.1: Supports for the structural parameters and parameters of the DGP, Smets and Wouters model.

Appendix C

In this appendix we show that failute to impose the uniqueness condition in identification could lead to large biases. For this purpose, 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 only 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 $(j,i), j =, 1,2,3; i = 1,\ldots,4$ elements of the matrix.

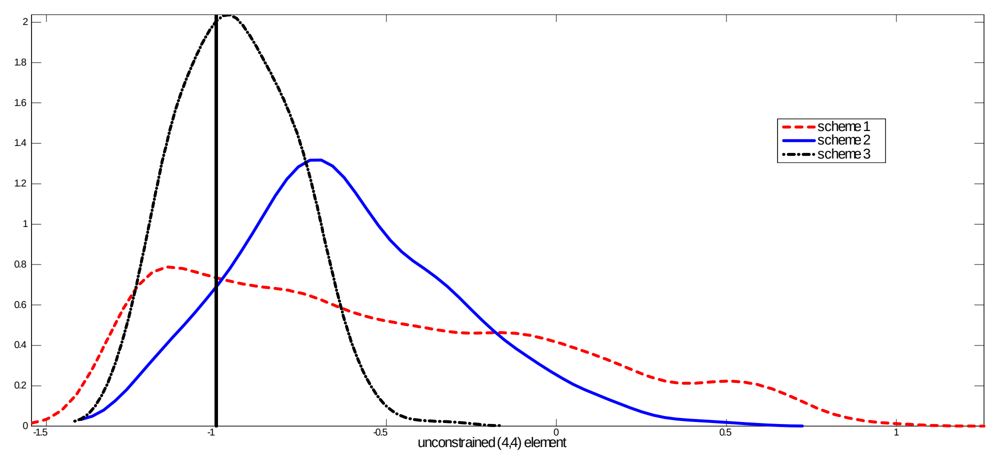

Figure C.1: Density of the response under different identification schemes. Scheme 1 sign restrictions, one shock; Scheme 2 sign plus uniqueness restrictions, on shock; Scheme 3 sign restrictions all shocks. Vertical bar: true value.

Figure C.1 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.