

# BUSINESS CYCLE MEASUREMENT WITH SOME THEORY <sup>\*</sup>

Fabio Canova  
ICREA-UPF, CREI, CREMeD, CEPR; <sup>†</sup>

Matthias Paustian  
Bank of England

First draft: November 2007;  
This draft July 21, 2011

## Abstract

A method to evaluate cyclical models not requiring knowledge of the DGP and the exact specification of the aggregate decision rules is proposed. 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 class of models is misspecified. The method is used 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.

---

<sup>\*</sup>We thank an anonymous referee, Heather Anderson, Luca Benati, Adrian Pagan, Frank Schorfheide for comments and the participants of seminars at CREI, ECB, EUI, Yale, University of Zurich, Bank of Hungary, Riksbank, Banque of France, Birbeck, 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.

<sup>†</sup>Corresponding author: Department of Economics, UPF, Ramon Trias Fargas 25-27, 08005, Barcelona (Spain), +3493521926, fabio.canova@upf.edu.

# 1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are nowadays regarded as the benchmark business cycle models for policy analysis and forecasting, both in academic and policy institutions. Their popularity is due to their attractive theoretical aspects and to the good forecasting performance 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. With a few notable exceptions (Del Negro and Schorfheide, 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 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

38 and Prescott, 1996) and pseudo-Bayesian predictive analysis (see e.g. Canova, 1995) to probabilistically  
39 evaluate the class, to discriminate among locally alternative data generating processes (DGP), and to provide  
40 information useful to respecify theoretical structures, if needed. Dedola and Neri (2007), Pappa (2009),  
41 Peersmann and Straub (2009), Lippi and Nobili (2010) among others, have used this methodology to answer  
42 interesting economic questions. What the paper provides is a formal presentation of the methodology, an  
43 assessment of its properties in simple experimental designs, and an application studying the role of rule-of-  
44 thumb consumers in generating realistic consumption responses to government expenditure shocks.

45 The analysis starts from a class of models which has an approximate state space representation once  
46 (log-)linearized around the steady state. We examine the dynamics of the endogenous variables in response  
47 to the disturbances for alternative members of the class using a variety of parameterizations and alternative  
48 specifications of non-essential (nuisance) aspects of the class. While magnitude restrictions depend on spec-  
49 ification details, the sign of the responses is much more robust to parameter and specification uncertainty.  
50 A subset of theoretically robust restrictions is then used to identify structural disturbances in the data and  
51 the dynamic responses of unrestricted variables are employed to evaluate the discrepancy between the class  
52 and the data or to select a member within the class.

53 The methodology has a number of advantages. First, it allows for misspecification in the structure to  
54 affect the likelihood function as long as it leaves the sign of the responses used for identification and testing  
55 unchanged. Thus, it is applicable to a richer class of problems than existing methods. Second, it can be  
56 employed to validate classes of models featuring more endogenous variables than shocks or rudimentarily  
57 specified dynamics. Third, by focusing shock identification and model testing on robust model-based quali-  
58 tative restrictions, the approach gives economic content to identification restrictions used in SVARs analyses  
59 and de-emphasizes the importance of a good calibration in testing the validity of a theory. Fourth, the  
60 procedure does not require optimization routines nor complex integration exercises and allows researchers  
61 to make identification and testing stronger or weaker depending on the needs of the analysis.

62 The approach can recover the sign of the impact response of unrestricted variables to the identified  
63 shocks, capture the qualitative features of the conditional dynamics, and exclude with high probability  
64 candidate DGPs in relevant designs. This occurs even when sample uncertainty exists, the empirical model  
65 is misspecified or the chosen class leaves important aspect of the DGP out. Finally, the approach can  
66 distinguish between sub-models in situations where standard approaches fail.

67 As an illustration, the methodology is used to gauge the frictions consistent with the observed transmission  
68 mechanism in the class of models with rule-of-thumb agents, suggested by Gali et al. (2007). The presence  
69 of a large number of non-optimizing consumers is insufficient to make consumption responses to government  
70 spending shocks positive. However, the robust restrictions the theory imposes can be employed to estimate  
71 the sign, the magnitude and the shape of consumption responses in the data. Since the share of non-  
72 optimizing agents needed to match the qualitative and quantitative features of conditional consumption  
73 dynamics in the data is unrealistically large, the validity of this class of models is called into question.

74 The rest of the paper is organized as follows. Section 2 illustrates the robust restrictions and the testable  
75 implications a class of models delivers. Section 3 describes the testing methodology. Section 4 studies the  
76 properties of the procedure. Section 5 evaluates a class of business cycle models. Section 6 concludes.

## 77 **2 From the theory to the data**

78 To illustrate the fundamental restrictions a theoretical structure imposes on the data and the nature of the  
79 testing exercise, the class of New-Keynesian models without capital, employed e.g. by Erceg et. al. (2000),  
80 Rabanal and Rubio-Ramirez (2005) among others, is considered.

81 The equilibrium conditions, with variables in log-deviations from the steady state, are in table 1.a. (T.1)  
82 is an Euler equation, (T.2) is a wage Phillips curve, (T.3) is a price Phillips curve, (T.4) is a Taylor rule,  
83 (T.5) defines the real wage and equation (T.6) is a production function. The economy is driven by four  
84 mutually uncorrelated, zero mean disturbances. The productivity shock  $e_t^z$  and the preference shock  $e_t^b$  have

85 autocorrelation coefficients  $\rho_z$  and  $\rho_b$ , respectively. The monetary shock  $e_t^R$  and the markup shock  $e_t^\mu$  are  
86 iid. The standard deviations of the innovations are  $(\sigma_z, \sigma_b, \sigma_R, \sigma_\mu)$ .

87 The goal is to derive restrictions which are robust to parameter variations, independent of the specification  
88 of nuisance features, and common to the sub-models in the class to identify shocks in the data and to test  
89 the validity of the class; and restrictions which are robust to parameter variations, independent of the  
90 specification of nuisance features but different across sub-models to select members of the class.

91 The structure represented in (T.1)-(T.6) is labeled M. The sub-models of interest are: a flexible price,  
92 sticky wage model ( $\zeta_p = 0$ ) (labelled M1); a sticky price, flexible wage model ( $\zeta_w = 0$ ) (labelled M2); a model  
93 with no indexation ( $\mu_p = 0, \mu_w = 0$ ) (labelled M3); a model with infinitely elastic labor supply ( $\sigma_l = 0$ )  
94 (labelled M4). Nuisance features in the class are the specification of habit and of nominal rigidities. In the  
95 basic specification, habit is additive and Calvo lotteries are used. As an alternative, multiplicative habit  
96 (labelled N1) and quadratic adjustment costs to prices and wages (labelled N2) are considered.

97 To obtain robust restrictions, a uniform distribution over an interval is specified for each structural  
98 parameter, chosen to be large enough to include theoretically reasonable values - see third column of Table  
99 1.b. For example, the interval for the risk aversion coefficient contains the values used in the calibration  
100 literature (typically 1 or 2) and the higher values employed in the asset pricing literature (see e.g. Bansal  
101 and Yaron, 2004), while the intervals for stickiness and indexation parameters include, roughly, the universe  
102 of possible values considered in the literature. While the interval for each parameter is independently and  
103 subjectively selected, in line with standard prior predictive analysis (see e.g. Geisser, 1980 or Kadane, 1980),  
104 one could make the ranges correlated and data based using the approach of Del Negro and Schorfheide (2008).  
105 The former approach is preferable here since it provides information about the range of possible outcomes  
106 the model can produce, prior to the use of any data. A large number of parameter vectors is drawn from  
107 these intervals, impulse responses are computed for each draw and pointwise 90 percent response intervals  
108 are extracted. 90 percent intervals trade-off two opposing forces: the desire to make the analysis as robust

109 as possible (which would suggest choosing large intervals); the awareness that, if the class is misspecified, no  
110 restriction will hold with probability one (which would suggest choosing small intervals).

## 111 **2.1 The restrictions**

112 Figure 1 shows the range of dynamic outcomes for the nominal rate, the real wage, price inflation rate,  
113 output, and hours for model M in response to monetary shocks. The magnitude of the responses depends  
114 on the parameterization. The sign of several dynamic responses is also fragile: the zero line is often included  
115 in the 90 percent interval at medium and long horizons. The sign of impact responses is instead robust: the  
116 impact interval for the nominal rate is positive; those for output, inflation and hours are negative.

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

121 Many impact responses have robust signs, both across sub-models and choices of nuisance features. For  
122 example, positive markup shocks increase production costs for any of the specifications and parameteri-  
123 zations, making production, the real wage and employment contract and inflation and the nominal rate  
124 increase. To test the validity of this class one could use, e.g., the restrictions that markup shocks produce  
125 on nominal rate, inflation, output and real wages to identify these disturbances in the data and then exam-  
126 ine whether the hours impact response interval is negative, as theory predicts. How many restrictions are  
127 used to identify and how many to test is question dependent. More identification restrictions avoid shocks  
128 confusion (for example, if only restrictions on output and inflation are used, markup and technology shocks  
129 are indistinguishable). More restrictions at the testing stage make the validation exercise sharper.

130 The impact response of the real wage to monetary disturbances is of interest since the sign of the  
131 interval differs for sub-models in the class featuring alternative nominal frictions. In sub-model M1 (flexible

132 prices and sticky wages), workers are off their labor supply schedule and from the firm's labor demand  
133 schedule,  $w_t = -\frac{\alpha}{1-\alpha}y_t$ , making real wages positively comove contemporaneously with monetary shocks. In  
134 sub-model M2 (sticky prices, flexible wages), workers are on their labor supply schedule and, on impact,  
135  $w_t = \left(\frac{\sigma_c}{1-h} + \frac{\sigma_l}{1-\alpha}\right)y_t$ , so that real wages are instantaneously negatively related to monetary shocks. Thus,  
136 to contrast sticky wages vs. sticky prices in the data, one could identify monetary shocks using the robust  
137 restrictions that the theory imposes on all variables but real wages and then examine whether real wages  
138 instantaneously fall or increase. Clearly, for testing to be meaningful, real wages need to be correctly  
139 measured, but such a problem is not specific to the approach proposed here.

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

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

### 148 **3 The mechanics of the evaluation approach**

149 The approach presumes that current business cycle models are still too stylized and feature too many black-  
150 box frictions to be taken seriously, even as an approximation to part of the DGP of the actual data (a  
151 point also made by Chari et al., 2009). This misspecification need not vanish by adding measurement errors  
152 or tagging artificial dynamics to the model, making standard measures of fit inadequate. By focusing on  
153 fundamental features of the propagation of shocks and using robust implications to distinguish alternatives,

154 the methodology sidesteps potential misspecification problems. To formally describe the approach, let

$$F(w_t^s(\alpha_0(\theta), \alpha_1(\theta)) | \epsilon_t, g, \mathcal{M}) \equiv F^s(\theta) \quad (1)$$

155 be a set of continuous model-based functions, computable conditional on the structural disturbances  $\epsilon_t$ ,  
 156 using models in the class  $\mathcal{M}$ , featuring the nuisance aspects  $g$ .  $F^s(\theta)$  could include impulse responses,  
 157 conditional cross correlations, distributions of conditional turning points, etc., and depends on the model-  
 158 produced series  $w_t^s$  via the coefficients of VAR representation of the decision rules, where  $\alpha_0(\theta)$  is the matrix  
 159 of contemporaneous coefficients,  $\alpha_1(\theta)$  the matrix of lagged coefficients and  $\theta$  the structural parameters. Let

$$F(w_t(\alpha_0, \alpha_1) | u_t) \equiv F(\alpha_0, \alpha_1) \quad (2)$$

160 be the corresponding set of data-based functions, conditional on the reduced form shocks  $u_t$ , where  $\alpha_0, \alpha_1$  are  
 161 the contemporaneous and lagged parameters of the VAR representation of the data. Both  $\theta$  and  $\alpha_0, \alpha_1$  are  
 162 treated as random variables. As it will be clear, identification and sampling variability make  $\alpha_0, \alpha_1$  random.  
 163 The class  $\mathcal{M}$  is assumed to be broad enough to include sub-models with interesting economic features. The  
 164 nuisance features  $g$  are not of direct interest but may affect the time series properties of  $w_t^s$ . The class  $\mathcal{M}$  is  
 165 misspecified in the sense that even if there exists a  $\theta_0$  such that  $\alpha_0 = \alpha_0(\theta_0)$  or  $\alpha_1 = \alpha_1(\theta_0)$ ,  $w_t^s(\theta_0) \neq w_t$ .  
 166 Thus, important aspects of the data (such as shocks, frictions or variables) may be left out of the class.

167 Among all possible  $F^s(\theta)$  functions, attention is restricted to the subset  $\tilde{F}^s(\theta)$  which are robust to  
 168 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  
 169 for shock identification and the  $J_2 \times 1$  vector  $\tilde{F}_2^s(\theta) \subset \tilde{F}^s(\theta)$  for evaluation purposes,  $\tilde{F}_1^s(\theta) \neq \tilde{F}_2^s(\theta)$ .  $\tilde{F}^s(\theta)$   
 170 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  $F^s$ ;  $\theta_l, \theta_u$  are the  
 171 upper and lower range of economically reasonable parameter values and the above holds for all interesting  
 172 specifications of  $g$ .  $\tilde{F}_1^s(\theta)$  must hold for all  $\mathcal{M}_j \in \mathcal{M}$ , while depending on what it is tested,  $\tilde{F}_2^s(\theta)$  may contain



173 functions whose sign does not depend on the sub-model (if generic fit is evaluated) or depends on  $\mathcal{M}_j$  (if  
 174 sub-models are compared). The economic question dictates what  $\tilde{F}_1^s(\theta)$  and  $\tilde{F}_2^s(\theta)$  will be.

175 To compute  $\tilde{F}^s(\theta)$ , one can follow Canova (1995), draw  $\theta$  from some prior distribution, solve the model,  
 176 and store  $F^s(\theta)$  at every draw. With the ordered output, one can then extract a credible interval and check  
 177 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.  
 178 To make sure that  $\tilde{F}_1^s(\theta)$  holds in the data, the covariance matrix of the reduced form shocks  $\Sigma_u$  is rotated  
 179 until  $\text{sgn}F(w_{1t}^s(\alpha_0(\theta), \alpha_1(\theta))|\epsilon_t, g, \mathcal{M}) = \text{sgn}F(w_{1t}(\alpha_0, \alpha_1)|u_t)$  where  $A_0A_0' = \Sigma_u$ ,  $\alpha_0 = A_0H$ ,  $HH' = I$ ,  
 180  $\alpha_1 = \alpha_0^{-1}A_1$ , where both  $A_1$  and  $\Sigma_u$  are drawn their empirical based distribution, and  $w_{1t}$  is the subset of  
 181  $w_t$  over which restrictions are imposed. An algorithm to efficiently generate  $H$  is provided by Rubio et al.  
 182 (2010). There maybe many, one or no  $H$  with the required characteristics. If no  $H$  exists, one can impose  
 183 the restrictions on another subset of  $w_{1t}$ , if available, or use another set of  $\tilde{F}_1^s(\theta)$ . If all interesting options  
 184 are exhausted and still no  $H$  is found, one can stop the evaluation process - the robust restrictions that the  
 185 class of models impose have no counterpart in the data. When  $k = 1, 2, \dots, K$   $H$  matrices are found, all the  
 186 generated  $(\alpha_0, \alpha_1)$  are stored.

187 Model evaluation then consists in probabilistic statements concerning the features of  $\tilde{F}_2(w_{2t}(\alpha_0, \alpha_1)|u_t)$ .  
 188 For example, one can compute the probability that  $\text{sgn}\tilde{F}_2(w_{2t}(\alpha_0, \alpha_1)|u_t) - \text{sgn}\tilde{F}_2^s(w_{2t}^s(\alpha_0(\theta), \alpha_1(\theta))|\epsilon_t, g, \mathcal{M}) =$   
 189  $0$  and  $w_{2t} \neq w_{1t}$  is a subset of  $w_t$ . Alternatively, one could compute the degree of overlap between the dis-  
 190 tribution of  $\tilde{F}_2^s(\theta)$  and of  $\tilde{F}_2(\alpha_0, \alpha_1)$ , where the distributions are obtained using the random draws of  $\theta$  and  
 191 of  $(\alpha_0, \alpha_1)$  obtained in the previous steps. If only one  $H$  is available,  $A_1$  and  $\Sigma_u$  are fixed at their sample  
 192 point estimate, one useful summary statistics is the probability that  $\tilde{F}_2^s(\theta) \leq \tilde{F}_2(\alpha_0, \alpha_1)$  where  $\theta$  are drawn  
 193 from  $[\theta_l, \theta_u]$ . Simple graphical devices, such as plots of the 90% bands in theory and in the data, could also  
 194 give a good idea of the likelihood of the restrictions.

195 To select among candidates the probability that  $\text{sgn}\tilde{F}_2(w_{2t}(\alpha_0, \alpha_1)|u_t) - \text{sgn}\tilde{F}_2^s(w_{2t}^s(\alpha_0(\theta), \alpha_1(\theta))|\epsilon_t, g, \mathcal{M}_j) =$   
 196  $0$  for each  $\mathcal{M}_j$  could be constructed and the sub- model with the highest probability chosen. Alternatively,

197 one could plot credible intervals for the sub-models of interest and take the one where the overlap with the  
198 theory is largest.

### 199 **3.1 Discussion**

200 The sign of the responses is used to derive robust constraints for two reasons: theory does not impose robust  
201 magnitude restrictions; and even if it did, magnitude restrictions need not hold in the data if the class of  
202 models is misspecified. Typically, impact restrictions are of interest, since as shown in section 2, the sign of  
203 the responses at longer horizons is generally not robust. When informational delays are present in theory,  
204 restrictions at longer horizons could be considered. Conditional functions, such as impulse responses, are  
205 preferred since they are more informative than unconditional moments about the features of  $\mathcal{M}$ .

206 The methodology is flexible and can be adapted to the need of the analysis. In fact, the identification  
207 process may involve more or less restrictions and one or more disturbances can be considered. Since standard  
208 rank and order conditions are not applicable to our case, how minimal this set of restrictions must be is  
209 generally unknown. Some indications on to proceed in practice are provided in the next section. Contrary  
210 to traditional practices, the identification restrictions are explicitly derived from a class of models and only  
211 robust constraints are considered. Thus, the procedure relies only on generic conditional dynamics and  
212 refrains from conditioning on a member of the class or on its parameterization.

213 The evaluation process is similar to the one employed in computational experiments where some moments  
214 are used to calibrate the structural parameters and others to check the goodness of the theory. Here a subset  
215 of the robust sign restrictions are employed to identify structural disturbances; the signs (and the shapes) of  
216 the dynamic responses of unrestricted variables are used to check the quality of the model's approximation to  
217 the data or to select a sub-model in the class. Two aspects are different: qualitative rather than quantitative  
218 restrictions are employed here at both stages; the evaluation process is probabilistic and takes into account  
219 both identification and sampling uncertainty.

220 Researchers are often concerned with the relative likelihood of sub-models in a class differing in terms  
221 of microfoundations, frictions, or functional forms. While the likelihood function need not be informative  
222 about these differences, our approach can, whenever sub-models differ in the sign (or the shape) of certain  
223 responses. For example, it is well known that sticky and flexible price versions of the same class of model  
224 produce different signs for the instantaneous response of hours to technology shocks. Once restrictions which  
225 are common to the two sub-models are used to identify technological disturbances, the response of hours  
226 can be used to discriminate the two theories. If sub-models differ in a number of implications, a weighted  
227 average of the relevant probabilities can be used to select the sub-model with the smaller discrepancy with  
228 the data. Candidate sub-models could be nested and or non-nested: the method works in both setups.

229 The approach compares favorably to existing methods for at least three reasons. First, the use of robust  
230 identification and testing restrictions shields researchers from model and parameter misspecification. Clearly,  
231 one cannot rule out the possibility that some type of misspecification changes the sign of key impulse  
232 responses; but qualitative restrictions on the sign of conditional moments tend to hold across many forms  
233 of misspecification. Second, the computational burden is smaller than the one involved in classical or  
234 Bayesian Likelihood-based evaluation techniques. Distributions of outcomes in theory are obtained when  
235 robust restrictions are sought; distributions of data outputs are obtained during the identification process  
236 and both require simple Monte Carlo exercises. Finally, the statistics one constructs can help to respecify  
237 the class, if the match with the data is unsatisfactory. For example, shape differences may suggest what type  
238 of amplification mechanism may be missing and sign differences the frictions that need to be introduced.

### 239 **3.2 The relationship with the literature**

240 The methodology is related to early work by Canova, Finn and Pagan, (1994) and Canova (1995), and to the  
241 recent strand of literature identifying VAR disturbances using sign restrictions (see Canova and De Nicolò',  
242 2002, or Uhlig, 2005). It is also related to Del Negro and Schorfheide (2004) and (2009), who use the data

243 generated by a cyclical model as a prior for reduced form VARs. Two differences set the approaches apart:  
244 the analysis here is conditional on a general class, rather than on a single model; qualitative rather than  
245 quantitative restrictions are used. This focus allows generic forms of model misspecification to be present  
246 and vastly extends the range of structures for which model evaluation becomes possible.

247 Corradi and Swanson (2007) developed a procedure to test misspecified models. Their approach is con-  
248 siderably more complicated, requires knowledge of the DGP and is not necessarily informative about the  
249 economic reasons for the discrepancy between the model and the data. Fukac and Pagan (2010) suggest to  
250 evaluate business cycle models using limited information methods but consider quantitative restrictions on  
251 single equations of the model while the focus here is on qualitative implications induced by certain distur-  
252 bances. Finally, Chari, et. al. (2007) evaluate business cycle models using reduced form "wedges". Relative  
253 to their work, a structural conditional approach and probabilistic measures of fit for model comparison exer-  
254 cises are employed. The emphasis on model evaluation techniques which do not employ statistical measures  
255 of fit is also present in Kocherlakota (2007), who shows that when the available candidates are all misspecified  
256 the best fitting model is not necessarily the more accurate for policy and inferential exercises.

## 257 **4 The evaluation procedure in controlled experiments**

258 To examine the properties of the procedure in realistic settings, either the small scale class of models described  
259 in section 2, or the larger scale version employed by Smets and Wouters (2003) are used as experimental  
260 DGPs. The analysis proceeds in two steps: in the first the properties of the procedure are investigated in  
261 population; in the second sampling and specification uncertainty are added to the setup.

### 262 **4.1 Population analysis**

263 Starting with the class of section 2, the flexible price, sticky wage sub-model M1 is selected as the DGP.  
264 The parameters used in simulating "pseudo-actual" data are the fourth column of table 1.b and similar to

265 the estimates of Rabanal and Rubio-Ramirez (2005). The researcher knows (T.1)-(T.6) and its solution,  
 266 meaning that both the model dynamics  $A_1$  and the covariance matrix of the reduced form errors  $\Sigma_u$  are  
 267 known. We ask whether the responses of the real wage can be recovered with high probability employing  
 268 different subsets of the robust restrictions, in alternative VAR systems, and identifying shocks either jointly  
 269 or separately. The matrix of impact coefficients is obtained as follows: i) a large number of normal matrices  
 270 with zero mean, unitary variance is drawn; ii) the QR decomposition is used to construct impact responses  
 271 as  $\alpha_0 = S*Q$ , where  $SS' = \Sigma$ ; iii) the responses satisfying the required restrictions are kept. To make results  
 272 stable, draws are made until 10000 candidates satisfying the restrictions are found. Thus here,  $\tilde{F}(\alpha_0, \alpha_1)$   
 273 reflects only identification but not sampling uncertainty.

#### 274 4.1.1 Can we recover the true model?

275 In the baseline case, the empirical model includes 5 variables: the nominal rate, output, inflation, hours  
 276 and the real wage. Since the economy features 4 structural shocks, a measurement error is attached to the  
 277 law of motion of the real wage when simulating data. Disturbances are identified (a) jointly, using robust  
 278 impact restrictions on all variables but the real wage; (b) jointly, using robust impact restrictions on all  
 279 variables but hours and the real wage; (c) individually, the markup shock; (d) individually, the monetary  
 280 shock. In (c) and (d), robust impact restrictions on all variables but the real wage are used. In addition  
 281 to the basic DGP, setups where either the standard deviation of monetary shocks or the standard deviation  
 282 of the markup shocks is 10 times larger are examined, and for each configuration, the four experiments are  
 283 repeated. Table 3 reports the percentage of correctly signed impact real wage responses.

284 The procedure recognizes the qualitative features of the DGP with high probability, in the ideal conditions  
 285 considered here. Two features of table 3 deserve attention. First, the number of shocks identified seems to  
 286 matter in some cases. For instance, in a 5 variable VAR and when a large standard deviation for markup  
 287 shocks is assumed, moving from identification scheme (d), which imposes restrictions only on responses to

288 monetary shocks, to identification scheme (a), which restricts responses to four structural shocks, raises the  
289 fraction of correctly signed responses to monetary shocks by 3 percentage points. In general, the benefit from  
290 identifying additional shocks when the economic interest is only in one particular structural shock depends  
291 on the DGP and seems to be larger when the variability of the shocks is more heterogeneous.

292 Second, as in Paustian (2007), the relative strength of the shock signal matters. For instance, when  
293 the standard deviation of the monetary shock increases tenfold, the fraction of correctly identified real wage  
294 responses to monetary shocks rises from about 72% to about 90% under identification scheme (d). Conversely,  
295 if the relative strength of the monetary shock signal is reduced, by increasing the standard deviation of the  
296 markup shock tenfold, the fraction of correctly signed responses to monetary shocks falls from roughly 74%  
297 to roughly 52%, again under identification scheme (d). On the other hand, the real wage effects of markup  
298 and taste shocks are easy to measure because their signal is relatively strong, making conclusions largely  
299 independent of the number of restrictions used and the number of shocks identified.

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

305 The same conclusions hold when hours is dropped from the VAR. A 4 variable VAR is fundamentally  
306 different from a 5 variable VAR since, in the latter, a state variable is missing - the observed real wage  
307 is a contaminated signal of the true one. Ravenna (2007) and Chari et. al. (2008) indicated that such  
308 an omission may be dangerous for inference if standard structural VARs are estimated. When robust sign  
309 restrictions on the impact response are used for identification, such an omission is less crucial.

#### 310 4.1.2 Can we exclude alternative models?

311 As Table 2 shows, a sticky price, flexible wage sub-model (M2) and a flexible price, sticky wage sub-model  
312 (M1) are local to each other as far as the sign of impact responses is concerned. The procedure can recover  
313 the sign of the real wage response to monetary shocks well when M1 is the DGP. Would the answer change  
314 if M2 and the parameterization listed in the last column of table 1 characterizes the DGP? Can the sign of  
315 the impact responses of the real wage to monetary shocks uncover the correct DGP with high probability?

316 The answer is positive. In the three experiments considered (identifying all shocks using the impact  
317 restrictions on output, inflation, hours and the nominal rate; identifying monetary, taste and supply shocks  
318 using impact restrictions on output, inflation and the nominal rate; and identifying only monetary shocks)  
319 the percentage of incorrectly recognized cases ranges between 0.4 and 1.3 percent. Could this conclusion  
320 be due to the selection of the parameters of the DGP? To examine this possibility, two other experiments  
321 are considered. First, the standard deviation of either the monetary or the markup shock is increased by  
322 a factor of ten. The conclusions are broadly unchanged: the fraction of impact real wage responses to  
323 monetary shocks that is incorrectly signed never exceeds 8 percent. Second, the parameters are randomly  
324 and uniformly drawn from the intervals shown in table 1.b. - in this case, 200 parameter vectors are drawn,  
325 setting  $\theta_w = 0$  for every draw, and for each vector, 10000 identification matrices are considered. When only  
326 monetary shocks are identified, the sign of the impact real wage response is incorrectly identified, on average,  
327 3.21 percent of the times. Thus, the exact parameterization has little influence on the results.

328 Why is the procedure successful in both capturing the DGP and in excluding local sub-models as potential  
329 data generators? While the range of impact real wage responses to monetary shocks obtained randomizing  
330 the parameters of the DGP in M1 and M2 is relatively large, the degree of overlap of the distribution of  
331 responses is minimal. Thus, one can tell apart the two sub-models with high probability because theory has  
332 sharp and alternative implications for the real wage responses to monetary shocks. The answer would be  
333 different if the implications of different sub-models were more muddled. For example, the response of the

334 real wage to technology shocks in M2 is not robust and the percentage of incorrect cases exceeds 25 percent  
335 under some identification configurations. Hence, only robust restrictions should be used for testing purposes.

336 These results are interesting also from a different perspective. Canova and Sala (2009) and Iskrev (2007)  
337 showed that classical econometric approaches have difficulties in separating sticky price and sticky wage  
338 models, because the distance function constructed using dynamic responses or the likelihood function are  
339 flat in the parameters controlling price and wage stickiness. Del Negro and Schorfheide (2008) report similar  
340 difficulties when Bayesian methods are used. The semi-parametric approach described here, which does not  
341 require structural parameter estimation, can potentially resolve the issue.

#### 342 **4.1.3 Summarizing the shape of the dynamic responses**

343 So far the sign of the impact response of a variable left unrestricted in the identification process is used to test  
344 the propagation mechanism of a sub-model. For many purposes this restricted focus is sufficient: business  
345 cycle theories do not typically have robust implications for the magnitude or the persistence of the responses  
346 to shocks. At times, however, the shape of the dynamic responses may be of interest. Alternatively, one may  
347 want to extend the testing to multiple horizons (if robust restrictions exist) and ask, for example, whether  
348 there exists a location measure that reasonably approximates, say, certain conditional multipliers.

349 Figure 2 plots the median of the set of identified real wage responses to shocks, horizon by horizon, and  
350 the true real wage responses in the basic setup, case (a) of table 3. The median is a good measure of the  
351 impact response of real wages to all shocks, both in a qualitative and in a quantitative sense. It also captures  
352 the sign of the dynamics well, but it is an imperfect estimator of the magnitude of the conditional real wage  
353 dynamics, at least as far as the responses to monetary shocks are concerned. Relative to other location  
354 measures, it is slightly better than the average response and very similar to the trimmed mean (computed  
355 dropping the top and the bottom 25 percent of the responses).

356 Fry and Pagan (2007) criticized the practice of using the median of the distribution as a location measure



357 when structural disturbances are identified with sign restrictions. Since the median at each horizon may be  
358 obtained from different candidate draws, identified shocks may be correlated. As an alternative, they suggest  
359 to use the single identification matrix that comes closest to producing the median impulse response for all  
360 variables. The correlation among identified shocks, computed using the median, ranges from 0.59 to 0.89 in  
361 absolute value depending on the experimental design. Therefore, Fry and Pagan's concern seems legitimate.  
362 However, as figure 2 shows, the alternative median is not a uniformly superior summary measure and its  
363 correlation with the true disturbances is generally low.

364 Several exercises were conducted to check the performance of the median in other experimental designs.  
365 The results suggest that (i) identifying more shocks or increasing the strength of the variance signal improves  
366 its performance; (ii) the dimensionality of the VAR is irrelevant for the dynamic properties of the median;  
367 and (iii) using model M1 or M2 as the DGP leaves the conclusions unchanged.

## 368 **4.2 Does sampling uncertainty matter?**

369 The ideal conditions considered so far are useful to understand the properties of the procedure but unlikely  
370 to hold in practice. What happens if the autoregressive parameters  $A_1$  and the covariance matrix of the  
371 shocks  $\Sigma_u$  are estimated prior to the identification exercise?

372 To capture estimation uncertainty, 200 replications of each experiment previously run are considered. In  
373 each replication, data is simulated, keeping the parameters fixed, and drawing shocks (and measurement  
374 error) from iid normal distributions with zero mean and standard deviations, as reported in table 1.b.  
375 Samples with 80, 160 and 500 points are considered. For each replication, a BVAR is estimated with a close  
376 to non-informative conjugate Normal-Wishart prior . An arbitrary fixed lag length is chosen because it is  
377 typical to do so in practice even though it adds misspecification - the decision rules imply that a VAR( $\infty$ )  
378 should be used. What happens if the lag length is optimally selected with BIC is also considered. The  
379 joint posterior of the dynamic parameters  $A_1$ , the covariance matrix  $\Sigma_u$ , and the identification matrices  $H$

380 is sampled until 2000 draws satisfying the restrictions are found for each replication. Table 4 reports the  
381 median value across replications of the probability that the impact response of the real wage to monetary  
382 shocks has the correct sign. Here the DGP is a sticky wage, flexible price model with one measurement  
383 error; a BVAR with the nominal rate, output, inflation, hours, and the real wage is estimated and shocks  
384 are identified imposing sign restrictions on the impact responses of the nominal rate, output, inflation and  
385 hours. Additional statistics for this experiment are in the accompanying materials (Appendix A) <sup>1</sup>.

386 Three features of table 4 stand out. First, sample uncertainty is small relative to identification uncertainty  
387 (see Kilian and Murphy, 2009, for related evidence) and the recognition probabilities do not clearly increase  
388 with the sample size, for each lag length. Second, changing the lag length of the VAR has little consequences  
389 on the outcomes. Since the same patterns are present when the lag length of the VAR is selected with BIC,  
390 none of the problems highlighted by Chari, et al. (2008) appear to be present here. Third, the number of  
391 shocks which are identified has minor consequences on the quality of the outcomes.

392 All other conclusions obtained in population hold also here. For example, the number of variables included  
393 in the VAR has little effect on the conclusions, and changing the variability of shocks produces the same  
394 results found in population. The DGP can be recognized and local sub-models can be excluded with high  
395 probability by looking at the impact response of the real wage to monetary shocks. Finally, the performance  
396 of the median, as a summary measure for the true responses, is broadly unaffected.

### 397 **4.3 Using the wrong model for inference**

398 We have argued that misspecification is generically less of a problem for the approach. To show that this  
399 is indeed the case, the procedure is next applied to a class of models which leaves out important aspects of  
400 the true DGP. For that purpose, data is generated from a version of the Smets and Wouters (SW) (2003)  
401 class of models and used to test the validity of the restrictions imposed by the class of models of section  
402 2. The smaller class has less shocks (investment specific, labor supply and government expenditure shocks

---

<sup>1</sup>Supplementary materials are available at JME in Science Direct.

403 are missing) than the SW class and the costs of adjusting investment and production frictions (fixed costs  
404 and variable capacity utilization) are disregarded. Since these differences are problematic for likelihood  
405 based methods, it is interesting to examine how large are the distortions that the approach would produce.  
406 The log-linearized optimality conditions, the parameter intervals used to derive robust restrictions and the  
407 parameters of the DGP are in the accompanying materials (Appendix B).

408 To begin with, it is useful check what robust restrictions the SW class imposes on output, inflation, the  
409 nominal rate, real wages and hours for each of the seven disturbances of the class. Table 5 reports the signs  
410 of the 90 percent impact response intervals. Interestingly, the sign of the intervals in responses to markup,  
411 monetary, taste and TFP disturbances are the same as in the basic model (compare with table 2) and robust  
412 across interesting sub-models. Thus, inference would not be necessarily distorted if a class of models which  
413 leaves out shocks and frictions present in the DGP is used to derive robust restrictions.

414 Table 5 also indicates that these restrictions alone may not be sufficient to uniquely obtain these four  
415 disturbances. In fact, in a five variable VAR, identified shocks may capture, in principle, any of the seven  
416 true structural shocks. For example, taste shocks could be confused with government expenditure shocks  
417 (four of the five signs are identical and for the fifth some confusion is possible), while markup and technology  
418 shocks may reflect investment specific shocks. To check the extent of the problem, the proportion of correctly  
419 signed real wage responses to shocks in population is computed. Some contamination is present, but it is  
420 remarkably small. For example, when markup, monetary, taste and technology shocks are identified using 16  
421 impact restrictions, the probabilities of correctly signing the impact real wage response are 98.1, 98.7, 90.7  
422 and 98.8, respectively. When only three shocks are identified using 12 impact restrictions, the probabilities  
423 are 98.6 for supply shocks, 99.5 for monetary shocks and 91.0 for taste shocks.

424 Since theory offers no guideline on the number of shocks to be included in a class of models, how can one  
425 limit shock confusion? Shrewdly choosing the variables of the VAR helps. As the last row of table 5 shows,  
426 if the labor productivity-real wage gap is added and the nominal rate is dropped from the list of observables,

427 the seven shocks produce mutually exclusive patterns of signs on the contemporaneous responses of the  
428 variables of interest. Thus, shock confusion is unlikely even if the smaller class is used for inference.

#### 429 **4.4 Testing multiple restrictions**

430 With the SW DGP one can also illustrate how the use of multiple restrictions - some of which may not be  
431 directly of interest - can strengthen testing in relevant practical situations. For the class considered, the  
432 instantaneous response of hours is robustly negative to TFP shocks if some price rigidities are present and  
433 robustly positive to labor supply, investment and markup shocks, regardless of the extent of price rigidities.  
434 The first implication is typically evaluated in the empirical literature, but hardly anyone seems to care  
435 about the other implications of the theory. However, jointly imposing the four restrictions may give sharper  
436 answers when price rigidities are weak, even if the latter restrictions are not of interest. To show this, data  
437 is simulated from the SW class using the same parameters as before except that  $\zeta_p = 0.3$  and  $\mu_p = 0$ . The  
438 probabilities that the impact response of hours is negative in response to TFP shocks and that the impact  
439 response of hours is negative in response to TFP shocks and positive in response to investment, labor supply  
440 and markup shocks are then computed.

441 The former probability is 61 percent indicating that, when price stickiness is low, it is difficult to distin-  
442 guish presence or absence of price rigidities. This probability increases to 83 percent when the four restrictions  
443 are jointly imposed - the difference is due to rotations matrices that imply negative hours responses to TFP  
444 shocks but also negative hours responses to any of the other shocks. Thus, when the data does not speak  
445 loud about the question of interest, imposing a larger set of restrictions can sharpen inference.

#### 446 **4.5 Advice to the users**

447 The procedure has good properties in all the experiments. However, three ingredients are needed to give  
448 the methodology its best chance of success. First, it is important not to be too agnostic in the identification

449 process. Sign restrictions are weak and this makes identification uncertainty important (see Manski and  
450 Nagy, 1998 for a similar result in micro settings). Thus, it is generally easier to recognize the DGP when  
451 more variables are restricted, for a given number of identified shocks, or more shocks are identified. Since  
452 theoretical sign restrictions at horizons larger than the impact one are often whimsical, constraints on the  
453 dynamic responses should be avoided at the identification stage. Similarly, sharper answers can be obtained  
454 if a number of robust restrictions, some which are of interest, some which are not, are jointly tested.

455 The experiments also showed that credible intervals tend to be large - this is expected given that the  
456 methodology delivers partially identified empirical models (see Moon and Schorfheide, 2009). Nevertheless,  
457 the probabilistic summary statistics employed are informative about the features of the DGP, even when  
458 asymptotically-based standard normal tests are not. If one insists on using the latter, a sufficient number of  
459 restrictions and smaller confidence intervals should be employed at the inferential stage.

460 Second, estimation biases should be, when possible, reduced since they may compound with identification  
461 uncertainty. In the experiments, estimation biases were small, even in small samples, but this need not to  
462 be the case for every possible design. A loose but informative prior was sufficient to reduce them. Other  
463 approaches, such as Kilian (1999), may work as well.

464 Third, inference is very reliable when the analysis focuses on the dynamics induced by shocks with stronger  
465 relative variance signal. However, even when the shock signal is weak, systematic mistakes are absent. While  
466 pathological examples can always be constructed (see Paustian, 2007, or Fry and Pagan, 2007), relative  
467 variance differences become a serious problem only in extreme circumstances. When interesting shocks are  
468 suspected to generate a weak relative signal, it is recommended to employ plenty of identification restrictions  
469 and to consider a class of models with a sufficiently rich shock structure. These two conditions were sufficient  
470 to ensure a good performance in all experiments we ran.

471 If a small scale class of models is used in the analysis, the choice of variables to be included in the VAR  
472 should be guided not only by economic but also by identification considerations. If the shocks produce

473 mutually exclusive patterns of robust signs for the impulse responses of the selected variables in theory, it  
474 is unlikely that the identified shocks mix true shocks of different type, making aggregation issues (see e.g.  
475 Faust and Leeper, 1997) less important.

476 In theory disturbances often generate a unique pattern of impact responses for the endogenous variables.  
477 In practice responses are not restricted to satisfy this uniqueness condition. Thus, when a subset of the  
478 shocks is identified, it is possible that shocks disregarded in the analysis generate similar pattern of responses.  
479 This multiplicity has no reason to exist and may make inference weaker than it should. As shown in the  
480 accompanying materials (Appendix C), failure to impose the uniqueness condition in identification may lead  
481 researchers astray. Thus, unless all shocks are identified, the condition should always be imposed.

482 Finally, as section 4.3 has shown, misspecification of the class of models does not necessarily imply wrong  
483 inference. In addition, the class of models used to derive the restrictions need not have the same number  
484 of shocks as the empirical VAR. All that is required is that any shock omitted from the structural model,  
485 but potentially present in the data, is not isomorphic to the shocks of interest in terms of signs of impulse  
486 responses. Thus, there is no need to arbitrarily add ad-hoc shocks to the structural model to conduct  
487 inference and starting from a good fitting class is not a precondition for the methodology to be applied.

## 488 **5 An example**

489 Standard business cycle models find it difficult to reproduce the private consumption dynamics in response to  
490 government expenditure shocks generated by structural VARs (see e.g. Perotti, 2007). However, one should  
491 also be aware that the restrictions used in this literature are not explicitly derived from any theoretical  
492 specification that is used to interpret the results. Gali et al. (2007) have taken a standard New Keynesian  
493 model and argued that adding one particular friction (a portion of non-Ricardian consumers) can make the  
494 theory consistent with the VAR evidence. This section investigates three questions. First, does the Gali  
495 et al. class of models produce positive consumption responses to spending shocks with high probability?

496 Second, what do consumption responses in the data look like if robust theoretical sign restrictions are used to  
 497 identify government spending shocks? Third, what is the likelihood that this class has generated the data?

## 498 5.1 The class of models

499 The log-linearized optimality conditions are in Table 6.a. Equations (T.7)-(T.8) describe the dynamics of  
 500 Tobin's  $q$ , its relationship with investments  $i_t$ . The law of motion of capital is in equation (T.9). Equation  
 501 (T.10) is the Euler equation of optimizing agents. Consumption of the non-Ricardian agents,  $c_t^r$ , depends  
 502 on their labor income obtained from supplying  $n_t^r$  hours at wage  $w_t$ , net of paying taxes  $t_t^r$ , where  $\alpha$  is the  
 503 share of labor in production, as in equation (T.11). The labor supply schedule for each group is in equation  
 504 (T.12). Cost minimization implies (T.13) and (T.14), where  $mc_t$  is real marginal cost,  $e_t^z$  a total factor  
 505 productivity shock and  $r_t$  the rental rate of capital. Output is produced as in (T.15). (T.16) indicates  
 506 that output is absorbed by aggregate consumption  $c_t$ , investment  $i_t$  and government spending  $e_t^g$ , which  
 507 is random. The new Keynesian Phillips curve is in equation (T.17), where  $e_t^u$  is an iid markup shock,  $\mu_p$   
 508 parameterizes the degree of indexation,  $\kappa_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}$ , and  $\zeta_p$  is the Calvo probability of non-changing  
 509 prices. The monetary policy rule is in equation (T.18) and  $e_t^R$  a monetary policy shock. The government  
 510 budget constraint and the fiscal rule give equation (T.19), where  $b_t$  are real bonds. The fiscal rule is in  
 511 (T.20). In the aggregate,  $c_t = \lambda c_t^r + (1-\lambda)c_t^o$ ,  $n_t = \lambda n_t^r + (1-\lambda)n_t^o$ ,  $t_t = \lambda t_t^r + (1-\lambda)t_t^o$ ,  $\lambda$  is the share of  
 512 non-Ricardian agents (ROTC), and  $t_t^j = \frac{T_t^j - T^j}{Y}$ ,  $j = o, r$ .

## 513 5.2 Evaluating the friction in theory

514 The literature often presumes that this class of models produces instantaneously positive consumption re-  
 515 sponses to government spending shocks when the share of ROTCs is sufficiently large. Is this a robust  
 516 implication of the theory? To check this, parameters values are drawn uniformly from the intervals in the  
 517 third column of Table 6.b, except for  $\lambda$  which is fixed at different values on a grid. The first panel of Fig-

518 ure 3, which reports the percentage of draws in which instantaneous consumption responses to government  
 519 spending shocks are negative for different  $\lambda$ , shows that the percentage increases with the share of ROTC  
 520 but a large  $\lambda$  is insufficient to robustly produce the desired result. In fact, even when the majority of the  
 521 consumers are not optimizers there is a non-negligible probability that reasonable parameters configurations  
 522 induce instantaneous negative consumption responses. The first panel of figure 3 also shows that if a large  
 523 share of ROTC is combined with large price stickiness, the required result obtains. Thus, while a large  
 524 value of  $\lambda$  is necessary, it is by no means sufficient. It is only when both  $\lambda$  and  $\zeta_p$  exceed 0.8 that one can  
 525 confidently conclude (say, with at least 90 percent probability) that this class has the required feature.

### 526 **5.3 Deriving robust identification restrictions**

527 Structural parameters are drawn from the intervals presented in the third column of Table 6.b, setting  $\beta =$   
 528 0.99, endogenously calculating  $c_y, i_y$  using steady state conditions, and keeping only those draws producing  
 529 a determinate rational expectations equilibrium - indeterminacy may occur for certain combinations of  $\lambda$   
 530 and  $\zeta_p$ . The range for most of the parameters is the same as in the experiments of section 4. For the fiscal  
 531 parameters, large intervals centered around the values used in the literature are selected.

532 Table 7 presents the sign of the 90 percent impact response intervals of output growth, inflation, hours  
 533 growth, investment growth to the four shocks. The combination of signs these intervals display is sufficient  
 534 to mutually distinguish all disturbances. This would not be the case, for example, if the nominal interest  
 535 rate is used in place of inflation (markup and monetary policy shocks will have similar sign implications).  
 536 Interestingly, 15 of the 16 sign restrictions displayed in the table remain if a positive correlation in the  
 537 intervals for  $(\gamma_\pi, \gamma_y)$ , for  $(\mu_p, \zeta_p)$  and for  $(\phi_b, \phi_g)$  is allowed. Only the response of inflation to expenditure  
 538 shocks is signed with less precision (around 65 percent) when  $\gamma_\pi$  and  $\gamma_y$  are sufficiently positively correlated.  
 539 Thus, having uncorrelated or correlated intervals makes little difference for the restrictions one derives.

540 Prior to the testing exercise, it is useful to check in a controlled experimental design whether the approach



541 can distinguish situations with and without non-Ricardian consumers using the restrictions of Table 7. The  
542 simulation uses the parameter values presented in the last column of Table 6.b (which are the same as in  
543 Gali et al., 2007). It is assumed that the researcher observes data on output growth, inflation, hours growth,  
544 investment growth and consumption growth and that the population VAR representation of these variables  
545 is known. For illustration, two polar cases are considered: no ROTC,  $\lambda = 0$ ; a large portion of ROTC  
546  $\lambda = 0.8$ . In both cases,  $\zeta_p$  is set to 0.75 to make the practical distinction between the two setups empirically  
547 relevant. Do the restrictions present in Table 7 allow us to sign the impact consumption growth response  
548 to government spending shocks with high probability? Do the dynamic responses of consumption growth in  
549 the VAR and in theory look similar? It turns out that in 99.6 percent of the accepted draws consumption  
550 falls on impact when  $\lambda = 0$  and in 78.2 percent of the accepted draws consumption increase on impact when  
551  $\lambda = 0.8$ . Furthermore, the median response path of consumption growth tracks the true response almost  
552 perfectly in both cases (see second panel of figure 3). Hence, the method can detect both the sign of the  
553 impact consumption responses and the shape of its dynamic responses to spending shocks, if the class of  
554 models has generated the data and if model-based restrictions are employed to identify spending shocks.

#### 555 **5.4 Is the friction relevant?**

556 A BVAR with a loose Normal Inverted-Wishart prior is estimated using quarterly U.S. data from 1954:1 to  
557 2007:2 obtained from the FRED database. The lag length of the VAR is two as selected by BIC. The BVAR  
558 includes, together with government consumption expenditure, output growth, GDP inflation, the growth rate  
559 of hours worked in the nonfarm business sector, and the growth rates of private investment and of private  
560 consumption. Four shocks are identified, imposing the 16 impact restrictions appearing in Table 7. The  
561 joint posterior of the BVAR parameters and orthonormal matrices is sampled until 1000 draws satisfying  
562 the restrictions are found. Data based error bands thus reflect sampling and identification uncertainty

563 The third panel of Figure 3 presents the responses of consumption growth to government spending shocks

564 in the data. When model based robust restrictions are imposed, consumption growth instantaneously in-  
565 creases. The point estimate is 0.25 and it is statistically significant but there is considerable uncertainty con-  
566 cerning the magnitude of the instantaneous consumption multiplier to spending shock (it could be anywhere  
567 between 0.06 and 0.45). Thus, the instantaneous consumption responses to spending shocks are comparable  
568 to those found in the micro literature for tax shocks (see e.g. Broda and Parker, 2008) Moreover, the increase  
569 is very short lived and after one quarter the 68 percent band includes zero.

570 Is the class of models a good candidate to explain the consumption responses observed in the data?  
571 To answer this question, the third panel of Figure 3 superimposes the theoretical consumption responses  
572 obtained when  $\lambda = 0.8$  and  $\zeta_p = 0.75$  while allowing all other parameters to be random. Clearly, the profile of  
573 the distribution of the responses in theory and in the data is similar. Instantaneously, the median responses  
574 are very close. At short horizons the median of the two distributions have similar size and shape and the  
575 probability that the sign of the responses in theory and in the data is the same is 83 percent on impact and  
576 72 percent over 2 horizons. Thus, to match the sign and the shape of the consumption responses observed in  
577 the data, considerable price stickiness and an unrealistically large share of ROTC are needed. Since micro  
578 evidence suggests moderate price stickiness, these results call into serious question the use of this class for  
579 inference and policy analyses <sup>2</sup>.

## 580 **6 Summary and conclusions**

581 A new methodology to examine the validity of business cycle models and to discriminate sub-models is  
582 presented. The approach employs the flexibility of SVAR techniques against model misspecification, the  
583 insights of computational experiments, and pseudo-Bayesian predictive analysis to link models to the data.  
584 Probabilistic measures of fit, which are robust to misspecification of the class and effective in providing  
585 information useful to respecify the class, are used to evaluate the discrepancy of the theory.

---

<sup>2</sup>As noted by Gali et. al., a model with imperfectly competitive labor markets may help to lower the share of rule of thumb consumers required to generate a rise in consumption to spending shocks.

586 The starting point of the analysis is a class of models which has an approximate state space representation  
587 once (log-)linearized around their steady states. The dynamics in response to shocks for alternative members  
588 of the class are examined using a variety of parameterizations and for different specifications of nuisance  
589 features. A subset of the robust restrictions is used to identify structural disturbances; another subset is  
590 used to measure the discrepancy between the class and the data or to discriminate members of the class. In  
591 controlled experiments, the approach can recognize the qualitative features of DGP with high probability  
592 and can tell apart local sub-models. It also provides a good handle of the quantitative features of the DGP  
593 if identification restrictions are abundant and if the relative variance signal of the shock(s) one wishes to  
594 identify is sufficiently strong. The methodology is successful even when the VAR is misspecified relative to  
595 the aggregate decision rules and when sampling uncertainty is present.

596 The methodology is appealing in several respects. First, it can be used even when the true DGP is not a  
597 member of the class of models one considers as long as the restrictions employed for identification and testing  
598 are not affected by the misspecification. Second, it does not require the probabilistic structure to be fully  
599 specified to be operative. Third, it shields researchers against omitted variable biases and representation  
600 problems. Fourth, it can be adapted to the needs of the user and requires limited computer time.

601 Apart from the illustrative example of section 5, recent work by Dedola and Neri (2007), Pappa (2009)  
602 Peersmann and Straub (2009) Lippi and Nobili (2010) among others, indicate the potentials that the method-  
603 ology possesses, the type of information it provides, and the interaction between theory and empirical work  
604 it produces. One interesting extension worth pursuing is transforming the evaluation approach into an esti-  
605 mation procedure, where the initial ranges for the parameters are updated using information similar to the  
606 one presented in Section 5. This approach, which provides an indirect way for obtaining interval estimates  
607 of the parameters, could become a useful alternative to likelihood based estimation approaches when the  
608 objective function is flat in the parameters of interest.

## References

- 609
- 610 Bansal, R., Yaron, A., 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles.  
611 *Journal of Finance* LIX, 1481-1509.
- 612 Broda, C., Parker, J., 2008. The impact of the 2008 Tax Rebates on consumer spending: preliminary  
613 evidence, manuscript.
- 614 Canova, F., 1995. Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium  
615 Economies. *International Economic Review* 36, 477-501.
- 616 Canova, F., 2009. Bridging cyclical DSGE models and the data, manuscript.
- 617 Canova, F., De Nicoló, G., 2002. Money Matters for Business Cycle Fluctuations in the G7. *Journal of*  
618 *Monetary Economics* 49, 1131-1159.
- 619 Canova, F., Gambetti, L., 2010. Do expectations matter? The Great Moderation revisited. *American*  
620 *Economic Journals: Macroeconomics* 2, 185-205.
- 621 Canova, F., Sala, L., 2009. Back to square one: identification issues in DSGE models. *Journal of*  
622 *Monetary Economics* 56, 431-449.
- 623 Canova, F., Finn, M. and Pagan, A., 1994. Evaluating a RBC model. In Hargraves, C., (Ed.) *Non*  
624 *Stationary Time Series Analysis and Cointegration*. Oxford University Press, Oxford.
- 625 Chang, Y., Kim, S., Schorfheide, F., 2010. Financial Frictions, Aggregation and Lucas critique, NBER  
626 working paper 16401.
- 627 Chari, V.V., Kehoe, P., McGrattan, E., 2007. Business cycle accounting. *Econometrica* 75, 781-836.
- 628 Chari, V.V., Kehoe, P., McGrattan, E., 2008. Are structural VAR with long run restrictions useful for  
629 developing business cycle theory?. *Journal of Monetary Economics* 55, 1337-1352.
- 630 Chari, V.V., Kehoe, P., McGrattan, E., 2009. New Keynesian Models: Not yet useful for policy analyses.  
631 *American Economic Journals: Macroeconomics* 1, 781-836.
- 632 Corradi, V., Swanson, N., 2007. Evaluation of Dynamic Stochastic General Equilibrium Models Based

633 on Distributional Comparison of Simulated and Historical Data. *Journal of Econometrics* 136, 699-723.

634 Dedola, L., Neri, S., 2007. What does a technology shock do? A VAR analysis with model-based sign  
635 restrictions. *Journal of Monetary Economics* 54, 512-549.

636 Del Negro, M., Schorfheide, F., 2004. Priors from General Equilibrium Models for VARs. *International*  
637 *Economic Review* 95, 643-673.

638 Del Negro, M., Schorfheide, F., 2008. Forming Priors for DSGE models (and how it affects the assessment  
639 of nominal rigidities . *Journal of Monetary Economics* 55, 1191-1208.

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

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

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

646 Fry, R., Pagan, A., 2007. Some Issues in Using Sign Restrictions for Identifying Structural VARs. NCER  
647 Working Paper No 14.

648 Fukac, M., Pagan, A., 2010. Limited Information Estimation and Evaluation of DSGE models. *Journal*  
649 *of Applied Econometrics* 25, 55-70.

650 Gali, J., López-Salido, J. D., Vallés, J., 2007. Understanding the Effects of Government Spending on  
651 Consumption. *Journal of the European Economic Association* 5, 227-270.

652 Geisser, S., 1980. A Predictivist Primer. In Zellner, A. (Ed.), *Bayesian Analysis in Econometrics and*  
653 *Statistics*. North-Holland, Amsterdam, 363-381.

654 Kydland, F., Prescott, E., 1996. The Computational Experiment. *Journal of Economic Perspectives* 10,  
655 69-85.

656 Kadane, J. B., 1980. Predictive and Structural Methods for Eliciting Prior Distributions. In Zellner, A.

657 (Ed.) Bayesian Analysis in Econometrics and Statistics. North Holland, Amsterdam, 89-94.

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

660 Kocherlakota, N., 2007. Model fit and model selection. Federal Reserve Bank of St. Louis Review 89,  
661 439-450.

662 Iskrev, N., 2007. How much do we learn from the estimation of DSGE models? A case study of Identifi-  
663 cation Issues in DSGE models, manuscript.

664 Lippi, F., Nobili, A., 2010. Oil and the macroeconomy: A structural VAR analysis with sign restrictions.  
665 Forthcoming, Journal of the European Economic Association.

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

668 Moon, K., Schorfheide, F., 2009. A Bayesian Look at Partially Identified Models, NBER working paper  
669 14882.

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

672 Paustian, M., 2007. Assessing Sign Restrictions. B.E. Journals of Macroeconomics, Topics, 7(1), article  
673 23.

674 Perotti, R., 2007. In Search of a Transmission Mechanism of Fiscal Policy. NBER Macroeconomic Annual  
675 2007, 169-226.

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

678 Rabanal, P., Rubio-Ramirez, J., 2005. Comparing New Keynesian models of the business cycle: A  
679 Bayesian approach. Journal of Monetary Economics 52, 1151-1166.

680 Ravenna, F., 2007. Vector Autoregressions and Reduced Form Representations of Dynamic Stochastic

681 General Equilibrium Models. *Journal of Monetary Economics* 54, 2048-2064.

682 Rubio-Ramirez, J., Waggoner, D., Zha, T. , 2010. Structural Vector Autoregressions: Theory of Identi-  
683 fication and Algorithms for Inference. *Review of Economic Studies* 77, 665-696.

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

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

688

Table 1.a: The equations of the model

$e_t^b - \frac{\sigma_c}{1-h} (y_t - hy_{t-1}) =$	$E_t[e_{t+1}^b - \frac{\sigma_c}{1-h} (y_{t+1} - hy_{t+1})] + (R_t - E_t\pi_{t+1})$	(T.1)
$\pi_t^w - \mu_w\pi_{t-1} =$	$\kappa_w \left[ -(e_t^b - \frac{\sigma_c}{1-h} (y_t - hy_{t-1})) + \sigma_l N_t - w_t \right] + \beta(E_t\pi_{t+1}^w - \mu_w\pi_t)$	(T.2)
$\pi_t - \mu_p\pi_{t-1} =$	$\kappa_p [w_t + n_t - y_t + e_t^\mu] + \beta(E_t\pi_{t+1} - \mu_p\pi_t)$	(T.3)
$R_t =$	$\rho_R R_{t-1} + (1 - \rho_R) [\gamma_\pi \pi_t + \gamma_y y_t] + e_t^R$	(T.4)
$w_t =$	$w_{t-1} + \pi_t^w - \pi_t$	(T.5)
$y_t =$	$e_t^z + (1 - \alpha)N_t$	(T.6)

689

690 The endogenous variables are  $y_t$ : output;  $N_t$ : hours worked;  $R_t$ : nominal rate;  $w_t$ : real wage rate;  $\pi_t$ : price inflation  
691 rate;  $\pi_t^w$ : wage inflation rate. The disturbances are: technology shock ( $e_t^z = \rho_z e_{t-1}^z + u_t, u_t \sim N(0, \sigma_z^2)$ ); preference shock  
692 ( $e_t^b = \rho_b e_{t-1}^b + v_t, v_t \sim N(0, \sigma_b^2)$ ); monetary policy shock ( $e_t^R \sim N(0, \sigma_r^2)$ ); and price markup shock ( $e_t^\mu \sim N(0, \sigma_u^2)$ ). In  
693 equation (T.3)  $\kappa_p \equiv \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p} \frac{1-\alpha}{(1-\alpha+\alpha\epsilon)}$  and in equation (T.2)  $\kappa_w \equiv \frac{(1-\zeta_w)(1-\beta\zeta_w)}{\zeta_w(1+\varphi\sigma_l)}$ .

694

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

Parameter	Description	Support	DGP1	DGP2
$\beta$	Discount factor	0.99	0.99	0.99
$\epsilon$	Elasticity in goods bundler	[5.00, 7.00]	6	6
$\varphi$	Elasticity in labor bundler	[5.00, 7.00]	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

695



696

Table 2: Signs of the impact response intervals to shocks.

	Markup shocks							Monetary shocks						
	M	M1	M2	M3	M4	N1	N2	M	<b>M1</b>	<b>M2</b>	M3	M4	N1	N2
$R_t$	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$w_t$	-	-	-	-	-	-	-	?	+	-	?	?	?	?
$\pi_t$	+	+	+	+	+	+	+	-	-	-	-	-	-	-
$y_t$	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$n_t$	-	-	-	-	-	-	-	-	-	-	-	-	-	-

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

697

698 A '+' indicates that at least 90 percent of the impact response interval is positive; a '-' that at least 90 percent of the  
699 impact response interval is negative; a '?' a response interval which lies on both sides of the zero line. M is the general model;  
700 in M1  $\zeta_p = 0$ ; in M2  $\zeta_w = 0$ ; in M3  $\mu_p = 0$  and  $\mu_w = 0$ ; in M4  $\sigma_l = 0$ . In N1 habit is of multiplicative form and in N2 nominal  
701 rigidities are modelled with quadratic adjustment costs.

702

Table 3: Percentage of cases where the impact real wage response is correctly signed.

		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		99.9		99.8		99.9		99.9		100		100	
Monetary		73.1	78.6		72.6	93.1	90.1		90.2	55.3	65.2		52.2
Taste		98.3	97.9			99.1	99.3			96.3	94.9		
Technology		99.5				99.6				97			
Supply			99.8				99.9				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			78.9		78.1		94.4		90.4		66.2		64.3
Taste			98.7				99.5				94.2		
Supply			99.8	99.6			99.8	99.8			99.9	99.8	

703

704 The VAR includes output, real wages, hours, inflation and the nominal rate in the first panel and output, real wages,  
705 inflation and the nominal rate in the second panel. In case (a) output, inflation, nominal rate and hours are restricted and  
706 shocks are jointly identified; in case (b) output, nominal rate and inflation are restricted and a supply shock, a monetary and a  
707 markup shock are identified; in cases (c) and (d) output, inflation, nominal rate and hours are restricted and a markup (supply)  
708 or a monetary shock are separately identified. In the second panel the standard deviation of either the monetary shocks is set  
709 10 times larger. In the third panel the standard deviation of either the markup shocks is set 10 times larger.

710

Table 4: Percentage of correctly signed real wage impact response to monetary shocks.

711

	All identified			Monetary shocks identified		
	T=80	T=160	T=500	T=80	T=160	T=500
VAR(2)	72	73	73	72	71	71
VAR(4)	73	72	73	72	71	72
VAR(10)	72	74	74	72	71	72
BIC	72	73	72	70	71	73

712

Median value across 200 Monte Carlo replications. The DGP is a flexible price, sticky wage model and the VAR includes

713

output, real wages, hours, inflation and the nominal rate.  $p = 2, 4, 10$  is the lag length of the VAR. The row labelled "BIC"

714

reports probabilities computed when the lag length of the VAR is selected with BIC.

715

Table 5: Signs of the impact response intervals to shocks, Smets and Wouter class.

716

	Markup	Monetary	Taste	Technology	Investment	Labor supply	Government
$y_t$	+	+	+	+	?	+	+
$\pi_t$	-	+	+	-	-	-	?
$R_t$	-	-	+	-	?	-	+
$w_t$	+	?	?	?	?	-	?
$n_t$	+	+	+	-	?	+	+
LP-W gap $_t$	-	?	-	+	+	-	-

717

A '+' indicates that at least 90 percent of the impact response interval is positive; a '-' that at least 90 percent of the

718

impact response interval is negative; a '?' a response interval which lies on both sides of the zero line.

Table 6.a: The equations of the model

$q_t = \beta E_t q_{t+1} + [1 - \beta(1 - \delta)] E_t r_{t+1}^k - (R_t - E_t \pi_{t+1})$	(T.7)
$i_t - k_{t-1} = \eta q_t$	(T.8)
$k_t = (1 - \delta)k_{t-1} + \delta i_t$	(T.9)
$c_t^o = c_{t+1}^o - (R_t - E_t \pi_{t+1})$	(T.10)
$c_t^r = \frac{1-\alpha}{\mu c_y} (w_t + n_t^r) - \frac{1}{c_y} t_t^r$	(T.11)
$w_t = c_t^j + \sigma_t n_t^j \quad j = o, r$	(T.12)
$r_t = mc_t + e_t^z + (1 - \alpha)(n_t - k_{t-1})$	(T.13)
$w_t = mc_t + e_t^z - \alpha(n_t - k_{t-1})$	(T.14)
$y_t = e_t^z + (1 - \alpha)n_t + \alpha k_{t-1}$	(T.15)
$y_t = c_y c_t + i_y i_t + g_y e_t^g$	(T.16)
$\pi_t - \mu_p \pi_{t-1} = \kappa_p (mc_t + e_t^u) + \beta (E_t \pi_{t+1} - \mu_p \pi_t)$	(T.17)
$R_t = \rho_R R_{t-1} + (1 - \rho_R)(\gamma_\pi \pi_t + \gamma_y y_t) + e_t^R$	(T.18)
$b_t = \frac{1}{\beta} [(1 - \phi_b) b_{t-1} + (1 - \phi_g) e_t^g]$	(T.19)
$t_t = \phi_b b_{t-1} + \phi_g e_t^g$	(T.20)

721 The disturbances are: technology shock ( $e_t^z = \rho_z e_{t-1}^z + u_t, u_t \sim N(0, \sigma_z^2)$ ); government spending shock ( $e_t^g = \rho_g e_{t-1}^g +$   
722  $v_t, v_t \sim N(0, \sigma_g^2)$ ); monetary policy shock ( $e_t^R \sim N(0, \sigma_R^2)$ ); and price markup shock ( $e_t^\mu \sim N(0, \sigma_u^2)$ ). The compound parameters  
723 in equation (T.17) is defined as:  $\kappa_p \equiv \frac{(1-\zeta_p)(1-\beta\zeta_p)}{\zeta_p}$ .

Table 6.b: Supports for the structural parameters and DGP used in the experiments.

Parameter	Description	Support	DGP
$\lambda$	Share of ROTC	[0.00,0.90]	0, 0.80
$\sigma_l$	Wage elasticity to hours	[0.00,1.00]	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
$\zeta_p$	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	[0.00,0.10]	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
$\phi_q$	Fiscal rule response to expenditure	[0.05,0.15]	0.1
$\rho_g$	AR(1) parameter government 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
$\sigma_R$	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

726

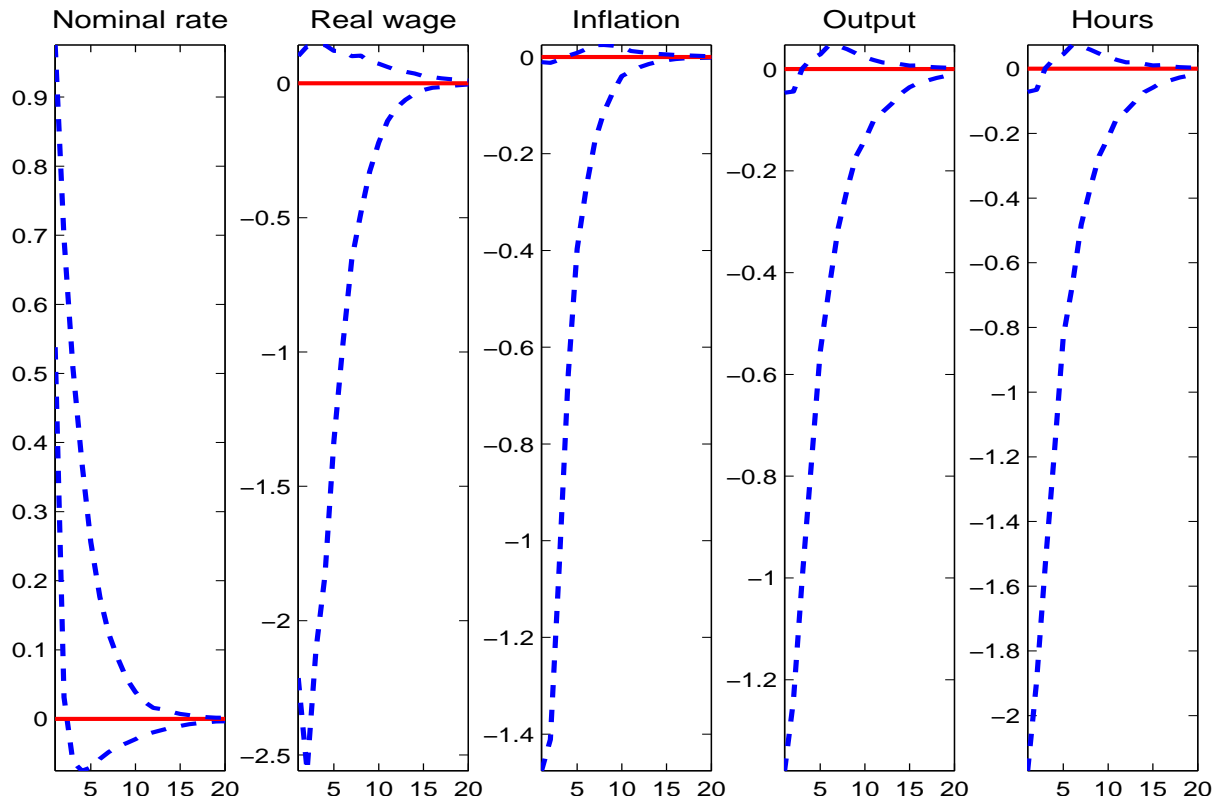
Table 7: Signs of the impact response intervals to shocks.

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

727

728 A '+' indicates that at least 90 percent of the impact response interval is positive; a '-' that at least 90 percent of the  
729 impact response interval is negative; a '?' a response interval which lies on both sides of the zero line. 10000 parameter vectors  
730 are drawn from the intervals in table 6

731



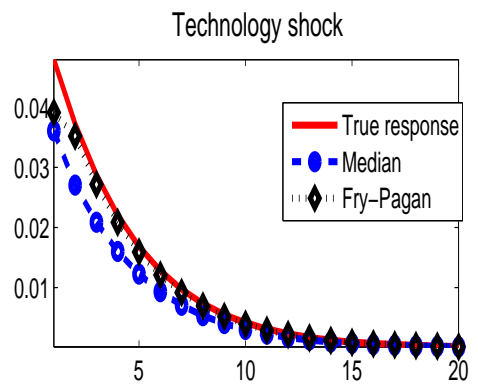
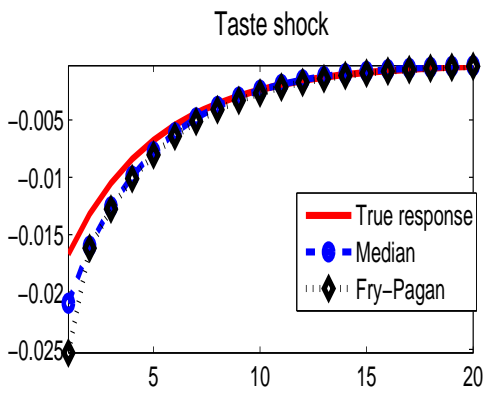
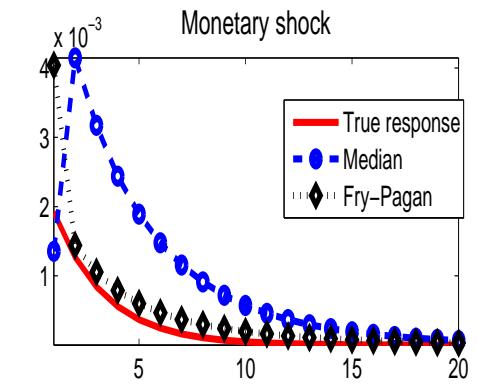
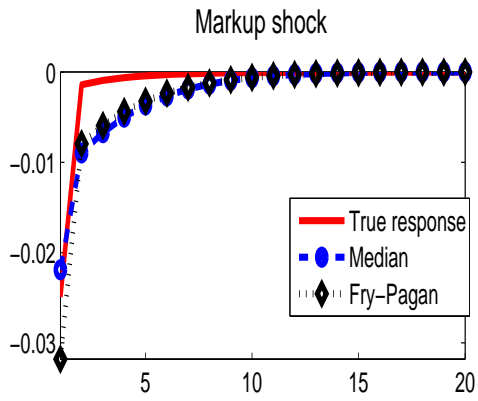
732

Figure 1: Pointwise 90 percent response intervals to monetary shocks. Model M.

733

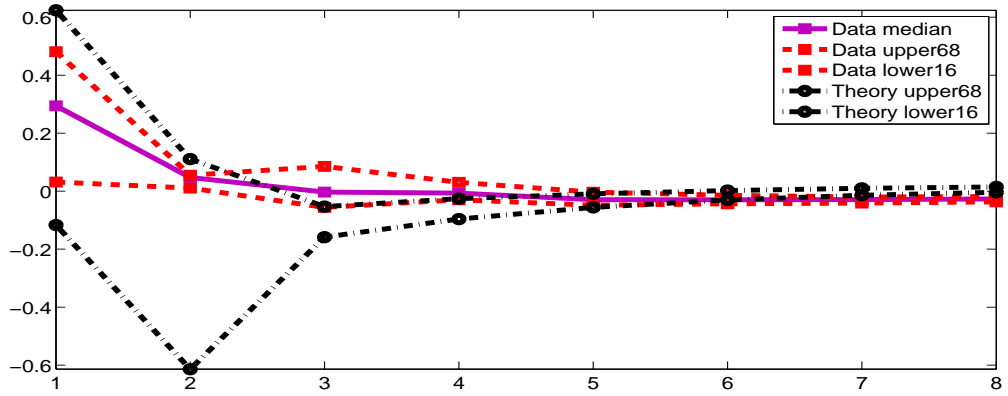
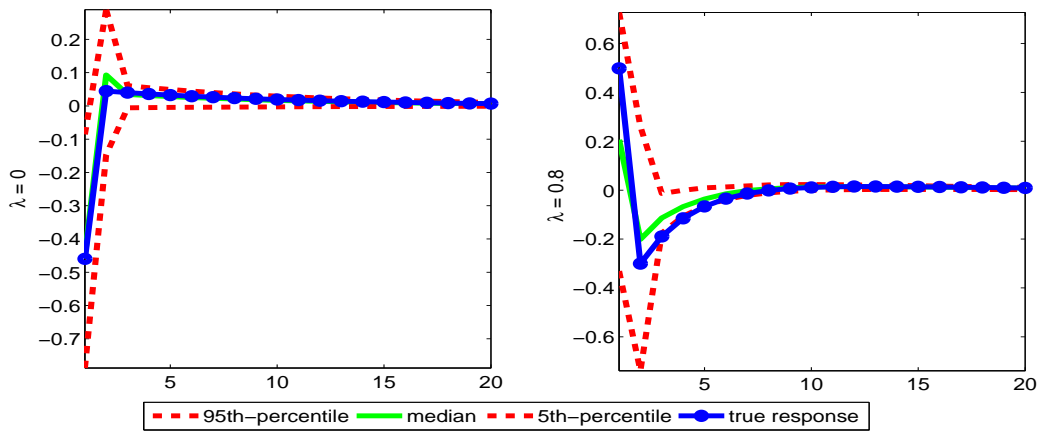
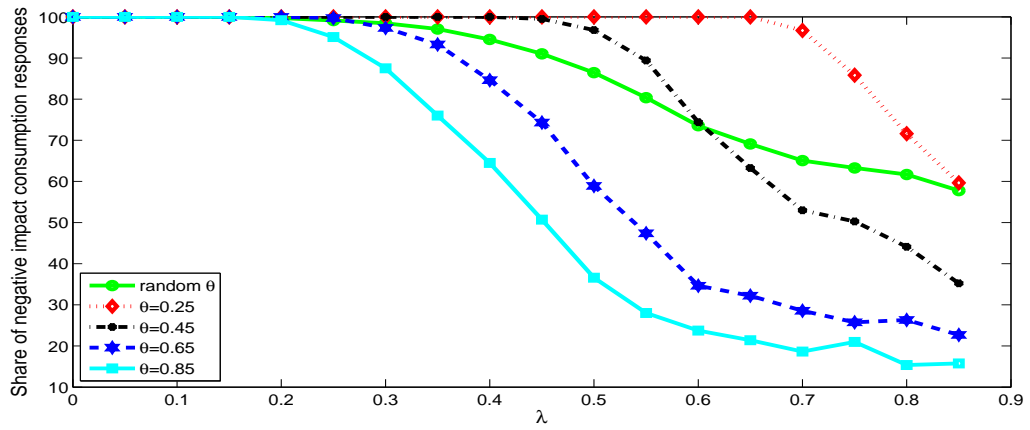


734



735

736



738 Figure 3: Consumption responses to government spending shocks. First panel theory; second panel simulated  
 739 data; third panel actual data.