Validating Monetary DSGE models through VARs

Fabio Canova
Universitat Pompeu Fabra, University of Southampton and CEPR

Very Preliminary, Please do not quote

* This paper grew out of the Panel Discussion of the workshop "DSGE Models and their use in monetary policy", held at the European Central Bank, June 5-6, 2001

July 25, 2001
1 Introduction

Two methodologies which have acquired a permanent status in the toolkit of macroeconomists in the last 15 years are Dynamic Stochastic General Equilibrium (DSGE) models and structural VAR. Although authors disagree on the exact microfoundations to be used in DSGE models and on the frictions needed to make them realistic they constitute, by and large, the framework in which new theories are developed and the improvement over existing ones examined. While the formulation of these models have been refined over the years and many realistic features have been added to the benchmark model popular, say, ten years ago, little progress has been made in designing and implementing tools to validate DSGE models against actual data. At the opposite extreme, VAR analyses eschew most of the detailed microstructure underlying DSGE and attempt to draw inference imposing only a minimal set of constraints to identify the dynamics in the data. Validation in these models is typically undertaken using a variety of statistical and economic procedures, which at times may result in circular arguments (see e.g. Uhlig (1999)), and a good deal of informal rules of thumbs.

Both theorists and policymakers are interested in the result of validation exercises for several reasons. Theorists typically want to know whether class of models is more appropriate than another to explain some phenomena or whether then introduction of an additional feature in a model improves its match with the data. Policymakers, on the other hand, need to make sure that a model represents a sufficiently good approximation to the data if they want to use them for policy purposes. To lend support to the various hypothesis one wants to entertain or to select a particular model which is a candidate for policy exercises validation techniques are simple, well understood, reproducible, and effective in comparing the discrepancy between the model to the data are clearly needed.

One reason for why validation methods for DSGE models are still at an infant stage, and informal measures of distances are typically used to compare the outcomes of the model and of the data, has been an earlier philosophical dispute about the intrinsic nature of an economic model. For example, it was claimed that existing statistical and econometric tools are appropriate to measure the fit to the data when the model represents the true DGP of the actual data (i.e. the error (discrepancy) is a random noise which has mean zero and it is uncorrelated with the model). However, when the model is only an approximation to the DGP of the data (and therefore misspecified in at least some dimension) econometric estimation and statistical testing becomes useless ("we reject a model known to be false"). In this situation the "computational experiment", as defined by e.g. Kydland and Prescott (1996), and its underlying calibration exercise, substitutes the standard set of econometric techniques used to validate models in the 70s and the 80s. Two points made by calibrators are well taken: the futility of applying statistical validation methods when the model is believed to be false; and the need of criterias which value economic discrepancy as opposed to statistical one. However, the insistence on rejecting econometric tools has made the
approach somewhat sterile (many exercises are now fancy back-of-the-envelope calculations) and led to a partial stall in the development of techniques which tells us how good a model is. Although the literature in the last ten years has attempted to incorporate to some extent the ideas put forward by followers of computational experiments within well established economic approaches (see e.g. Christiano and Eichenbaum (1992), Watson (1993), Canova (1994), DeJong, Ingram and Whitemann (1996) and Geweke (1999) among others), several problems still remain.

The task of this paper is to provide a new validation methodology for DSGE models that integrates "calibration" and VARs approaches and attempts to overcome some of the problems which are present with existing ones. The starting point of our analysis is a class of DSGE models which have an approximate (log)-linear representation around their steady states. We take seriously the objection that such models are at best approximations to the DGP of the actual data and even as approximations they may work only in some dimensions. We also take seriously the idea that economic validation is what matters for users of DSGE models and we design an evaluation procedure which takes these concerns into account. We are also sensitive to the objection that too little sensitivity analysis is typically performed on calibrated DSGE models - and therefore that the outcomes of the model may depend on particular and arbitrary parameter choices. For this reason, our approach searches for robust implications of the model, implications which hold regardless of the exact parametrization, of particular functional choices, certain policy rules or of some additional features that a researcher may want to include in the specification. Once robust implications are discovered, we used them as devices to identify structural shocks in VAR models. That is, we transform a reduced form VAR into a structural model by imposing a minimal set of robust restrictions derived from a DSGE model. We then use responses to identified shocks in the data and in the model to provide a metric to compare both qualitatively and quantitatively the dynamic outcomes. Based on the results obtained at this stage one can then proceed to respecify the model, in case the discrepancy is large, answer interesting economic questions or conduct policy analyses of some sort.

There are several advantages of our proposed approach: it simple; it is easily reproducible; it integrates calibration and VAR analyses in a framework where both become useful; it produces implications which are robust to the exact calibration of the DSGE model and to assumptions about certain functional forms; finally, and, contrary to several existing procedures, which are designed to be effective with one particular type of shock, it can used to compare models using dynamic response of different shocks.

We apply this methodology to the problem of evaluating two, by now standard, DGSE models of money - a limited participation economy and a sticky price monopolistic competitive economy - against the data. The use of monetary DSGE models has grown enormously in the last few years: a large portion of the theoretical literature works with variations of models in these two classes. Furthermore, these types of models are routinely used to analyze policy questions (Philips’ curve, welfare consequences of different policies, the relative
importance of monetary disturbances for real fluctuations, optimal monetary and taxation issues) both in academics and policy centers. Despite of this widespread use and interest, there have been surprisingly very few studies which examined how the two class of models perform on a common set of questions or their relative merits in explaining the data (see e.g. Tack (1996), Christiano, Eichenbaum and Evans (1997), Chari, Kehoe and McGrattan (2000)). Our exercise can therefore shed important light on the nature of these models and on their match with the data which can be useful to both theorists and policymakers.

Our results suggest that the both models are at odds with the data in several ways. First, for some data sets the robust restrictions that these model deliver have no counterpart in the data. In other words, fundamental implications of the models are qualitatively inconsistent with the data. Second, the qualitatively dynamics of both models are different than those in the data. For example, the limited participation model has hard time to replicate the sign of the responses of labor productivity following a monetary policy shock and the sticky price model generate dynamics for the slope of the term structure which are inconsistent with those present in the data and lack at times of liquidity effects. Third, quantitatively speaking, the persistence of the real effects in the sticky price model fall short of those present in the data and both models fail to capture the size of output effects in response to monetary policy disturbances. Also, the limited participation model appears to be more suited than the sticky price model to capture the effects of policy shocks on inflation. Fourth, while the two models generate dynamics in response to technology disturbances which are qualitatively and also quantitatively similar, the data constrained with the restrictions produced by the model generate responses which are fragile, in the sense that they depend on the data set used and to some extent on the set of restrictions we impose. Fifth, we show that two simple alteration suggested in the literature are unable to bring the responses of labor productivity in the limited participation model more in line with the data suggesting that more drastic modifications are needed to improve the dynamic fit of the model in this dimension. Finally, we show that the welfare implications of different monetary policy rules are very similar in the two models: in response to a policy shock a partial accomodative rule, which makes the nominal interest rate respond to real balances, is to be preferred when consumer’s welfare is considered. The relative ordering of Taylor and inflation targetting rule depend instead on the model.

The rest of the paper of the paper is organized as follows: the next section briefly review the literature on evaluating DSGE models and points out the weakness of each existing approach. Section 3 describes in details the suggested methodology. Section 4 presents the evaluating of the two DSGE models of money. Section 5 concludes.
2 The state of the art

The problem of validating DSGE models has attracted substantial interest for at least 10 years and several approaches have been suggested in the literature. In this section we briefly summarize the current state of the art, highlighting advantages and disadvantages of the various approaches and the problems that still remain unsolved.

The oldest approach available in the literature builds on work by Sargent (1978) and Altug (1985) and takes the DSGE model as restricted but probabilistically incomplete representation of the actual data. From this point of view, "structural estimation" and "structural evaluation" are feasible with standard statistical tools (for example, maximum likelihood) once the probabilistic structure of the model has been completed either arbitrarily adding dynamics (see Berkowitz, Diebold, Ohanian (1999)), measurement errors Ireland (1999) or shocks (e.g. Ingram, Savin and Whitemann (1994), Leeper and Sims (1994), Smets and Wouters (2001)). In some cases direct estimation of the structural relationships is attempted with the probabilistically incomplete setup of the model (see Ireland (2001)). Given that the modified representation is assumed to represent correctly the DGP of the actual data - at least under the null - statistical validation by standard goodness of fit (likelihood ratio, etc.) can be easily undertaken and cross equation restrictions typically provide a wealth of constraints which can be used to further test the model. Statistical testing has been recognized to provide little information about economic relationship. Therefore formal statistical methods are sometimes used in conjunction with informal ones, and the information concerning the "reasonableness" of moments, cross correlation functions, and impulse responses is used complement the information coming from statistical tests.

There are several problems connected with this approach. First, the way the probabilistic structure of the model is completed is arbitrary and leaves lots of room for non-comparability across studies. Second, both in the case where the probabilistic structure of the model is completed and where it is not, computational problems typically emerge (because the likelihood function has multiple peaks or flat areas, or because optimization routines are unable produce estimates when information about some parameters is "weak"). Third, parameter estimates often turn out to be unreasonable from an economic point of view and tend to lie on the boundary of the admissible parameter space. These last two problems taken together usually suggest that misspecification is still present and that statistical tests, which take the model to be correct under the null, may be inappropriate. Finally, and excluding some relevant exceptions, the economic implications of the estimates are not discussed, interesting economic exercises are not performed and statistical validation represents the end of the exercise.

Recently, and to avoid problems with parameters which are difficult to estimate, many authors have used a mixed approach, which selects a-priori some parameters and estimate the rest. This standard quick fix makes computations easier, since the restricted likelihood function is typically much better behaved, but may jeopardize the entire validation exercise,
since the distribution of estimated parameters depends in a non-trivial way on the values of the parameters which are fixed a-priori. Gregory and Smith (1991) showed, in the simple context of the equity premium model, that the distribution of estimates are skewed and non-normal when parameters which are fixed a-priori are not consistent estimates of true ones. Hence inference about functions of these estimates, such as impulse responses, cross correlations or simple moments, are likely to give a misleading representation of the dynamics which one would obtain when all the parameters are jointly estimated. One recent example along these lines is Leeper and Zha (2000), who show that inference concerning the magnitude of the coefficients of a Taylor rule (in particular, the coefficient on the inflation gap) are not independent of the values assumed for certain parameters in other parts of the model.

A second strand of literature takes the view that a DSGE model is false even when the dynamic specification is enriched and/or shocks are added. That is, a model is only a stylized description which leaves out important features of the data. Adding dynamics or measurement error will not necessarily modify the fact that important features are left out. Given this point of view, estimation of parameters via the likelihood function is unlikely to be successful because of the large misspecification and even when this is possible parameter estimates are unlikely to be meaningful, making statistical validation tests uninformative (the model will be rejected no-matter what).

According to the most extreme version of view, standard econometrics should be dispensed of and economic questions should be answered using computational experiments similar to those undertaken in experimental and physical sciences (see e.g. Kydland and Prescott (1996)). Validation is implicit in the calibration of the experiment (the model fits some observations by construction): therefore the calibration of the model becomes the crucial step to make the outcomes credible. Additional informal exercises (back-of-the-envelope-calculation) may be performed to evaluate the ability of the model to reproduce stylized facts of the data and the credibility of the model to answer interesting questions is enhanced if more stylized facts are matched. Despite statements suggesting the contrary, computational experiments are designed and performed using a well defined statistical (method of moment) criteria, and evaluation is based on the choice of a specific loss function (which weighs only first moments) (see e.g. Canova (1994)). The process of calibrating a model to the data typically leaves a number of parameters unspecified and therefore induces some arbitrariness in the comparison. This and the fact that the evaluation process is done informally has drawn strong criticisms (see Hansen and Heckmann (1996)). In particular, the lack of measures of uncertainty attached to the outcomes of the model has been seen as detrimental to the establishment of secure fundations to the approach. Recently, however, Broze, Dridi and Renault (1999) have provided choice theoretical fundations to calibration exercises and shown there are conditions under which an informal evaluation may be the optimal thing to do.

A less extreme point of view, maintains the hypothesis that a DSGE model is a simplified and therefore false representation of the GDP data, but also suggests that it may provide
a reasonable representation for certain aspects of the data. Hence, instead of attempting to estimate/evaluate the model as if it provided a description of the likelihood of the events observed in the data, one should should be modestly concerned with the more limited information approach of estimating/evaluating the model on those specific aspects (statistics, moments) of the data it was designed for. Within this general point of view, several classes of approaches have emerged which differ in metric used to assess the closeness of the model to the data and in type of randomness allowed to exist. Watson (1993), for example, asks how much variability should a measurement error have so that a calibrated DGSE with this measurement error produced dynamics which quantitatively similar to those in the actual data. In this case, the metric for the comparison is given by the $R^2$ of the regression of the data on the model and no variability in the parameters is allowed. Christiano and Eichenbaum (1992), Feve and Langot (1994) and others use sampling variability of the actual data to provide GMM style estimate of the parameters of interest and goodness of fit tests for the additional moment restrictions not used in the estimation process. On the other hand, some authors have used sampling variability of model to construct simulated method of moment (SMM) estimates of the unknow parameters and confidence bands/probabilistic statements for the goodness of the model. In the literature there are examples where these exercises are conducted using the randomness of the exogenous processes (Gregory and Smith (1993)) or a realistic pametrisation of the uncertainty surrounding the selection of the parameters of the model (Canova and Marrinan (1996), Maffezzoni (2000)). More recently, approaches which take a symmetric view about the uncertainty surrounding both actual data and the model have emerged (see e.g. Canova and Denicolo (1995), DeJong, Ingram, Whitemann (1996), (2000)) and Bayesian methods have been developed to validation validate models in a framework where even the few moments used to estimate/ evaluate are misspecified (see Geweke (1999) and Schorfeide (2000)).

Several problems have been highlighted with this class of approaches. For example, Geweke (1999) has argued that there are logical inconsistencies in the view that the model is a good representation of only some dimensions of the data. He also suggests that these inconsistencies can be solved only if the DSGE is taken to be a representation for the moments of the data (not of the data or a selected group of time series). In addition, validation methods in these procedures are built on assumption that the model truely characterizes some the aspects of the data, an assumption which is probably as heroic as the one that the model is the correct GDP for the entire data set. Furthermore, the previously noted problem of calibrating some parameters and estimating others is more acute in these setups because a researcher has a degree of arbitrariness in selecting the features of the data which constitute the benchmark against which the model is evaluated. Moreover, most of the approaches rely on statistical measures which are only at best partially informative on economic issues of interest. Finally, if one takes the radical view that even the few moments under consideration may be misspecified, the estimation/ evaluation process becomes computationally difficult. Computational constraints may become binding and the methods turn out to be impractical.
or hard to deal with when parameter space is large (say, greater > 5) or number of time series under consideration is moderately large (greater than 3-4).

3 An alternative procedure

The approach we propose in this paper attempts to avoid some of the problems mentioned in the previous section. The basic starting point of the analysis is that DSGE models are too stylized to be taken seriously even as an approximation to the DGP of (part of) the actual data. Therefore, the methodology does not exclude the possibility that the model is misspecified. Our task will be to try to find a set of implications which are robust to the exact parametrization of the model, and use them to restrict the actual data. Once the data has been ”calibrated” to look like the model in some dimensions, we will attempt to validate the latter in other dimensions by qualitatively and quantitatively examining the mismatch between the model predictions and the data. The discrepancy between the model and the data is measured using economic (not statistical) concepts and the features of the internal propagation mechanism of the model in response to shocks are used to provide a metric to assess the difference between the model and the data. If the mismatch is found be not too large, policy analyses can be undertaken and answers to interesting questions provided. Otherwise one can proceed to respecify the model and repeat the exercise.

Our approach share similarities with both existing structural VAR and calibration methodologies. As in structural VARs we use a minimal set of restrictions to identify the object of interest in the analysis. Contrary to the standard practice, we derive these restrictions explicitly from a DGSE model instead of appealing to conventional ones, which are at times unreasonable and inconsistent with the dynamics of DGSE models (see Canova and Pina (2000)). Moreover, we look for restrictions which are uncontroversial from the point of view of the model. As in standard calibration exercise, we use some observations to make the data and the model look alike, but we reverse the common practice by using restrictions of the model to calibrate the data. Also, as it is done in that literature, we compare the economic outcomes of the model with those of the data and we do this both qualitatively and quantitatively. However, contrary to most of the analyses we do not compare reduced form implications of the model and the data, but instead examine directly the structural dynamics in response to shocks of both. In addition, while existing procedures typically allow, at most, the identification of one shock, our approach is flexible and permits the identification of every shock for which a robust characterization of the dynamics it produces can be found in the model. This allows a more thorough evaluation of the theoretical economy since we can simultaneously analyze the implication for the endogenous time series of different structural shocks (e.g. monetary, technology, fiscal, etc.). Finally, the method is straightforward, is computationally simple and reproducible.

The procedure involves six specific steps:
1. Find robust implications of a class of models.

2. Use implications in the data to identify shocks. Verify if restrictions are born out in the data.

3. Evaluate the performance of empirical/theoretical models qualitatively examining responses of endogenous variables to shocks.

4. Cross validate qualitatively across models

5. If answers to 3.-4. positive, and some policy analysis should be performed, continue the validation quantitatively.

6. Respecify the model or undertake policy analyses.

The first step of our procedure is explicitly designed to cope with the inherent arbitrariness of calibration procedures. We define an implication of the model to be robust if it holds independent of parametrization or the functional forms for the primitives. For example, if the unconditional covariations of output and the nominal interest rate are qualitatively similar when we vary the risk aversion parameter within a reasonable range of values, we call this implication robust. Robustness is not generic since many implications of model are sensitive to the exact parameterization employed and to specific features added or subtracted to the model. What we are looking for here is a set of implications which are representative of the class of models we want to evaluate. As we will see later on, robust implications typically takes the form of shape or sign restrictions. At times, there may be magnitude restrictions which may be robust. However, given that the model is misspecified, magnitude restrictions are unlikely to born out in the data. While both unconditional and conditional comovements of variables of the model can be used, we find it more useful to examine comovements conditional on different shocks, since most of the empirical analyses have predictions on the way endogenous variables react to identified structural shocks.

The second step involves making the model and the data look alike in certain dimensions. In particular, we would like to identify disturbances which produce dynamics which are qualitatively similar in the model and in the data. This amount to calibrating the data to the model. Clearly, it is possible that the restrictions derived by the model are not born in the data. In that case, one would either repeat the exercise imposing alternative but robust implications of model on the data, or, if all robust implications are examined and no shock in the data that produce dynamics that looks like those in the model yet found, stop the evaluation process.

The third step is similar to that employed in many evaluation procedure. Typically, some moments (statistics) are used to estimate/calibrate the parameters, others are used to check the performance of the model. We do the same here. We stress that at this stage the evaluation is qualitative, is intended to examine whether the dynamics of the model
and the data are broadly consistent and involves essentially the examination of the sign and the shape of dynamic responses of certain variables. For example, if a "supply" shock is identified by means of the joint responses of output and inflation, we would like to see whether the response of employment to this shock is qualitatively similar in the model and in the data. Alternatively, one may want to check whether the humped shape response of capital in response to supply shocks in the model is also found in the data.

At times one is not necessarily interested in the absolute performance of a model but instead may be interested in the relative comparison of models which differ in some relevant features, for example microfundations or economic frictions. These features can, at times, be nested (a model with capacity utilization or without) but most of the times they are not (e.g. monopolistic competitive vs. perfectly competitive markets). Once the first three steps of our evaluation procedure have been undertaken and none of the models immediately discarded, it is the possible to qualitatively compare models. For example, following Chari, Kehoe and McGrattan (2000) one may ask whether the restrictions imposed by model X are better than those of model Y in making output dynamics in response to monetary shocks looks more similar in the model and in the data, or whether the addition of a feature Z to the model enhance the match, given some robust identifying restrictions. Once again, we keep the comparison qualitative, at least at the beginning: many models can be discarded relative to others using simple qualitative implications such sign and shape features of the responses.

Next, it may be the case that a model need to be used to perform some policy exercises or to give answers to certain quantitative questions. For instance, one may be concerned with measuring the size of the real effects of monetary policy or the timing of the responses of certain variables to structural shocks, e.g. whether the timing of a peak response of a variable changes if different policy actions are taken. One example of the latter type is whether monetary policy rules of Taylor type produce quantitatively different responses of macrovariables than inflation rules. Alternatively, one may want to measure whether welfare losses due to distortions and frictions are larger in a model or another (as e.g. Smets and Wouters (2001). In all these cases standard Monte Carlo methods can be used to assess the significance of the differences and construct probabilities of interesting events (as e.g. Canova (1994)).

4 Evaluating two monetary models

To illustrate how the procedure works in a concrete example, we examine the performance of two standard DSGE models used to study the propagation of monetary shocks. The first is a version of the Limited Participation economy studied, among others, by Christiano, Eichenbaum and Evans (CEE) (1997), Cooley and Quadrini (1999) and Canova and Pina (1999). The second is a version of a monopolistic competitive model with sticky prices.
examined, among others by Rotemberg and Woodford (1997), King and Wolman (1999), Gali (1999) and Smets and Wouters (2001). Contrary to a large portion of the literature, we allow for capital accumulation in both models, since the measurement of short run effects in the data may be distorted if this feature is excluded. Furthermore, we use a general form for the utility function of the agents which nests specifications used in the literature. The two model are identical in several respects; they differ primarily in the mechanics of transmission of the disturbances. In the former model, monetary shocks have real effects because they alter the opportunity costs of hiring productive labor. In this model money has therefore effects because it alters the production possibility of the economy. In the latter, real effects obtain because monopolistic competitive firms accept to meet whatever level of aggregate expenditure is set in the economy at prices which are exogenously allowed to vary only in a predetermined and sluggish manner. Since the evaluations of models of this type has been partially carried out in the literature (e.g. Tack (1996), CEE (1997), Chari, Kehoe and McGrattan (2000), Leeper and Zha (2000), Kim (2000)) with alternative techniques, we have some benchmarks against which examine whether our evaluation method shed further light on the performance of the two classes of models.

4.1 Limited Participation model

In this economy there is a large number of consumers maximizing a utility function which values consumption and leisure of the form

$$\max_{\{c_t, n_t, k_{t+1}, I_t\}} E_0 \sum_{t} \beta^t \left( \frac{c_t^{\mu} (1 - n_t)^{1-\mu}}{\gamma} \right)$$

by choices of consumption ($c_t$), hours ($n_t$), capital $k_{t+1}$ and deposits ($I_t$) subject to three constraints: a cash-in-advance constraint, a budget constraint and a capital accumulation constraint of the forms:

$$c_t p_t \leq M_t - I_t + W_t n_t$$  \hspace{1cm} (2)

$$M_{t+1} = D_t + F_t + r_t p_t k_t + M_t - I_t + W_t n_t - c_t p_t - x_t p_t$$  \hspace{1cm} (3)

$$x_t = k_{t+1} - (1 - \delta) k_t + \frac{a}{2} \left( \frac{k_{t+1}}{k_t} - 1 \right)^2 k_t$$  \hspace{1cm} (4)

where $W_t n_t$ is nominal labor income, $r_t p_t k_t$ is nominal capital income, $D_t$ and $F_t$ are the profits obtained by owning the firms and the banks and $a$ is an adjustment cost parameter. Implicit in this formulation are three important facts: deposit are chosen before shocks are realized; consumers undertake the investments at the end of each period subject to an adjustment cost which is quadratic; labor income can be used to purchase consumption goods in the same period while all types of capital income can not be used for this purpose. Note also that, by construction, there are no costs of adjusting capital in the steady state.
There is a large number of competitive firms which hire capital and labor to produce an homogenous good. Firms face a working capital constraint in maximizing their profits which forces them to finance their wage bill before the receipts of the sales of the goods are obtained. To obtain working capital they borrow funds from the banks. Their problem is

$$\max_{\{k_t, n_t\}} F_t = (P_t v_t k_t^\alpha n_t^{1-\alpha} - P_t r_t k_t - (1 + R_t)W_t n_t)$$  \hspace{1cm} (5)$$

where $R_t$ is the nominal interest rate and $v_t$ is a productivity disturbance.

Banks receive deposits from consumers and lend them together with the new injection of money obtained from the monetary authority to the firms. The problem they face is

$$\max_{\{B_t\}} D_t = ((1 + R_t)(B_t - I_t))$$  \hspace{1cm} (6)$$

by choices of loans ($B_t$) subject to the constraint that $B_t = I_t + S_t$ where $M_{t+1} - M_t = S_t$ are new injections of money.

Finally, we assume that the monetary authority fix interest rates according to a Taylor-type rule of the form:

$$R_t = Y_t^{\omega_1} \pi_t^{\omega_2} \epsilon_t$$  \hspace{1cm} (7)$$

where $Y_t$ is output, $\pi_t = \frac{p_t}{p_{t-1}}$ is the gross inflation rate and $\epsilon_t$ is a monetary policy disturbance. Given this rule, the monetary authority stands ready to provide whatever amount of money is demanded by consumers at that nominal interest rate.

To summarize, the states of the economy are ($k_t, I_t, m_t$) where $m_t = \frac{M_{t+1}}{p_{t-1}}$; there are two types of shocks ($v_t, \epsilon_t$) and the model determines 12 endogenous variables ($c_t, n_t, Y_t, R_t, r_t, x_t, \pi_t, i_t = \frac{I_t}{p_t}, d_t = \frac{D_t}{p_t}, b_t = \frac{B_t}{p_t}, s_t = \frac{S_t}{p_t}, w_t = \frac{W_t}{p_t}$). In addition, it is possible to compute two other variables which we employ to evaluate the model: labor productivity ($LP_t = \frac{Y_t}{n_t}$) and the slope of the nominal term structure ($SL_t = R_t^\infty - R_t$), where $R_t^\infty$ is the long run nominal interest rate.

The responses of the endogenous variables to a contractionary monetary shock for the standard parameter values reported in table 1 are presented in the first column of figure 1: the nominal interest rate increases and this contracts employment and output via the working capital constraint. Since agents prefer to smooth consumption, the decrease in consumption is smaller than the decrease in output. Hence investment must decrease to maintain the resource contraint satisfied. Because of the adjustment costs, the disinvestment process is slow and inflation temporarily increases to insure that aggregate demand matches the aggregate supply of goods. Because of the cash-in-advance constraint the fall in consumption produces a fall in real balances and a liquidity effect is obtained. Finally, the slope of the term structure decreases while labor productivity temporary increases in response to the shock.

The responses of the endogenous variables to an expansionary technology disturbances are standard (see second column of figure 1): hours, output and investments increase following
the shock and consumption increases by a smaller amount (the shock is temporary and agents want to smooth consumption). The increase demand of labor from firms produces an increase in the demand for funds and given the policy rule, this induces an increase in the nominal interest rates. Inflation falls since the increase in aggregate supply is larger than the increase in aggregate demand and real balances increase to match the increase in consumption. Finally, labor productivity increases, as the increase in output is larger than the increase in employment, and the slope of the term structure falls.

4.2 Monopolistic competitive-sticky prices model

The sticky price (SP) model we consider to a large extent similar to the limited participation (LP) economy we have described above. Here we simply highlight differences when they emerge. There is a large number of identical consumers maximizing a utility function defined on consumption, leisure and real balances of the form

$$\max_{\{c_t, n_t, k_{t+1}, M_{t+1}\}} E_0 \sum_t \beta^t \left( \frac{c_t^\gamma (1 - n_t)^{1-\mu}}{\gamma} + \frac{\theta}{1-\nu} \left( \frac{M_{t+1}}{p_t} \right)^{1-\nu} \right)$$

by choices of consumption ($c_t$), hours ($n_t$), capital ($k_{t+1}$) and real balances ($m_{t+1}$) subject to the following sequence of budget constraints

$$M_{t+1} \leq r_t p_t k_t + W_t n_t + M_t + S_t + F_t - (x_t + c_t) p_t$$

and the capital accumulation constraint (4), where $F_t$ are the profits from owning the firms producing intermediate goods, $p_t = (f_0^1 p_t^{1-\rho} d_i)^{\frac{1}{1-\rho}}$; $c_t = (f_0^1 c_i^{1-\rho} d_i)^{\frac{1}{1-\rho}}$.

Intermediate firms are monopolistic competitive and set price according to a time dependent rule. Their problem is to minimize costs

$$\min_{\{k_i, n_i\}} (r_t k_{it} + W_t n_{it})$$

subject to the production function constraint $Y_{it} = v_i k_{it}^{\alpha} n_{it}^{1-\alpha}$. In choosing the price to charge for their goods, these firms maximize expected discounted profits

$$\max_{\{P_{t+j}\}} \frac{U''(c_{t+1})}{P_{t+1}} \eta^{\frac{1}{\rho}} F_{t+j}$$

subject to the demand function $\frac{Y_{it+j}}{P_{t+j}} = (\frac{P_t}{P_{t+j}})^{-\rho} j = 0, 1, 2, ...$, where $F_{t+j} = (p_{t+j} Y_{t+j} - MC_{it+j})$, $MC_{it+j}$ is the marginal cost of firm $i$ at time $t + j$; $\eta^{\frac{1}{\rho}}$ is the probability that a price set at time $t$ will still prevail at time $t + j$.

Given the pricing decision of firm, the aggregate price level evolves according to

$$p_t = (\eta p_{t-1}^{1-\rho} + (1-\eta)p_t^{1-\rho})^{\frac{1}{1-\rho}}$$
where \( \tilde{p}_t = \left( \frac{1}{\rho} \tilde{p}_t^{\lambda - \rho} - \rho \right)^{\frac{1}{1 - \rho}} \); \( \tilde{p}_t = \frac{\rho}{\rho - 1} \sum_{j=0}^{\infty} \frac{u'(c_{t+j})}{r_{t+j}} Y_{t+j} MC_{it+j} \) and \( \frac{\rho}{\rho - 1} \) is the steady state markup (the inverse of the steady state real marginal cost).

Finally, we assume that the monetary authority sets nominal interest rates according to (??) and that at that rate it stands ready to provide the amount of money consumers demand.

To summarize, the states of the economy are \((k_t, m_t)\); there are two types of shocks \((v_t, \varepsilon_t)\) and the model determines 10 endogenous variables \((c_t, n_t, Y_t, R_t, \pi_t, \pi_t, f_t = \frac{\Delta}{p_t}, s_t = \frac{S_t}{p_t}, w_t = \frac{W_t}{p_t})\). As with the previous model we also compute equilibrium paths for two additional variables: labor productivity \((Lp_t = \frac{Y_t}{n_t})\) and the slope of the nominal term structure \((SL_t = R_t^\infty - R_t)\).

Impulse responses following a contractionary monetary policy disturbance for the standard parametrization of the model reported in table 1 are presented in the first column of figure 2. It is very easy to check that the impact effects are qualitatively similar to the one presented in figure 1: the nominal interest rate temporarily increases following the shock and this temporarily contracts employment and output. Since the decline in consumption is smaller than the decline in output, investment decreases as well. Because of the adjustment costs, the disinvestment is slow and inflation temporarily increases to insure that aggregate demand equals the aggregate supply of goods. With the adopted parametrization real balances fall in response to the shock and a liquidity effect is obtained. Finally, the slope of the term structure decreases while labor productivity temporary increases. While the impact effects in figures 1 and 2 are qualitatively similar the persistence of the reponses is minimal with the SP model (as in Chari, Kehoe and McGrattan (2000)): the effect of monetary shocks on output, employment, productivity and consumption dies out in one period and only real balances and the capital stock slowly converge to their original steady state.

The responses to an expansionary technological disturbances (second column of figure 2) are also very similar to those obtained with the limited participation model: employment, output, capital and consumption all increase, the nominal interest rate increases following the expansion of output above its steady state, while inflation temporarily declines to maintain aggregate demand equal to aggregate supply. Quantitatively speaking, the increase in employment and output is much larger than the one obtained in the limited participation model while the response of real balances is substantially smaller. Finally, as a consequence of these differences, the increase in labor productivity in the SP model is smaller than the one obtained in the LP model but is hump shaped, while the decline in the slope of term structure is stronger than the one obtained in the LP economy.
4.3 Evaluation

4.3.1 Robust features

To examine the validity of the two prototype models we have described we first attempt to find some implications of the dynamic responses of variables to shocks which do not depend on the exact calibration of the parameters. We initially focus attention to the responses of the endogenous variables to monetary disturbances since these two models differ primarily in the way monetary disturbances are transmitted to the real economy. Later we discuss the restrictions and the implications that these models provide in response to technological disturbances.

Figures 3-4 report the pairwise cross correlation function of inflation, output, real balances, and the slope of the term structure in response to monetary disturbances for the LP and the SP economy, respectively. In figure 3 the first column reports the correlations for the basic LP model (whose responses are presented in figure 1), the second column the correlations when there are some adjustment costs (a=2), the third column correlations when the utility function is logarithmic in consumption and labor is indivisible; the fourth column the correlations when the Taylor rule is substituted by a rule where nominal interest rates react to real balances (with coefficient equal to 0.8), as in Leeper and Zha (2000), the fifth column reports correlations when the coefficient of relative risk aversion is large (and equal to 5) and the last column reports correlations when adjustment costs are extreme (a=30). In figure 4 the first column reports the correlations for the basic SP model (whose responses are in figure 2), the second column the correlations when prices are changed more often (%( = 0.25), the third column reports correlations when the elasticity of money demand is high (ε = 20); the fourth column the correlations when the Taylor rule is substituted by a rule where nominal interest rates react to real balances (with coefficient equal to 0.8), the fifth column correlations when the utility function is logarithmic in consumption and labor is indivisible and the last column reports correlations obtained when a money growth rule is employed in place of a Taylor rule.

Both figures indicate that the sign (and, in some cases even the magnitude) of these correlations is very similar across different parametrizations of the models. In particular, the LP economy generates negative cross correlation between inflation and output, and between inflation and real balances in response to monetary disturbances, while the correlation between output and real balances is positive throughout the range except, perhaps at one year lead with two specifications. Furthermore the cross correlation of the slope and output and of the slope and real balances is positive; while the cross correlation of inflation and the slope is negative in all the range. The SP economy, on the other hand, generates robust V shaped patterns: the one between inflation and output are positive contemporaneously and for lags of output and negative for leads of output. The correlation between inflation and real balances is negative everywhere, while the correlation between output and real balances is positive.
for lags of real balances and negative contemporaneouly and for leads of real balances. The 
cross correlation of the slope and output is S-shaped going from negative to positive values; 
the correlation of the slope with inflation is negative and the one between the slope and real 
balances is positive.

These sign patterns are present with all parametrizations we have considered except in 
the last columns of the two figures. In the last column of figure 3 the adjustments costs 
are so high that capital is constant; therefore these would be the cross correlations obtained 
in a model without capital. In the last column of figure 4 the shape of cross correlations 
functions are altered but notice that for the first three cross correlations contemporaneous 
sign restrictions are still the same. To avoid illegible figures we have omitted graphs of 
the cross correlations obtained with other specifications, including one where steady state 
inflation is high and one where agents’ utility value only consumption, but we stress that 
the qualitative features of the dynamics we consider here are identical in these alternative 
setups.

We conclude that the restrictions in the signs of the cross correlation function for all 
pairs of variables in response to monetary shocks are robust to variations of parameters 
within a reasonable range, to alternative policy rules and, partially, to choices concerning 
the functional forms for the primitives.

4.3.2 Identification of shocks

One could use some or all of these restrictions to characterize the dynamics induced by 
monetary shocks in the two models. For reasons which will become apparent later on, we 
will limit attention to a subset of these constraints and select the sign restrictions obtained 
on the cross correlation function of output, inflation and the slope for the LP model and 
on the cross correlation function of output, inflation and real balances in the SP model. 
One motivation for choosing two different sets of correlations for the two models is that 
technology and monetary disturbances imply different sets of sign restrictions for these three 
variables in the two models. We then run a five variable VAR with output, inflation, real 
balances, the slope of the term structure and labor productivity and identify as monetary 
those disturbances which produce the same sign restrictions on the cross correlation function 
of the three pairs of variables as in the models. In searching for shocks which satisfy the 
restrictions, we follow Canova and Denicolo (1999), whose approach is briefly summarized, 
as a reference in appendix A. Intuitively, the procedure systematically searches in the space 
of identifications to find ”structural” shocks which produce responses with the requires sign 
characteristics. Figures 5 and 6 report the three pairs of cross correlations in response 
to identified monetary disturbances for three different countries: the US, the UK and the 
EURO land. In all cases the sample covers quarterly data from 1980:1 to 1998:4. In figure 
5 we presents correlations obtained when the LP sign restrictions are imposed and figure 
6 presents correlations obtained when the SP sign restrictions are used. Both figures are
constructed imposing sign restrictions only on the contemporaneous elements of the cross correlation function.

Two features of the figures deserve comments. First, while the monetary shocks we derive satisfy the sign restrictions we have imposed, the shapes of the cross correlation functions we have extracted from the data are somewhat different from the theoretical ones. Second, LP sign restrictions fail to find any monetary shock with the UK data, while SP sign restrictions do not produce monetary shocks with Euroland data. That is, with these two data sets, we have found no combination of reduced form residuals which produces cross correlations for output, inflation and the slope (or real balances) with the required contemporaneous sign. Hence, the support for the LP model with UK data and for the SP with Euroland data is negligible. This lack of identifiability, which we consider the first important test of the two theories, however, it is not peculiar to these two cases. In fact, it generalizes to both models and all data sets had we been more demanding with our identification restrictions. For example, if in addition to contemporaneous restrictions we had imposed sign restrictions on the first lead and the first lag of the cross correlation function of the three pairs of variables, we would not be able to identify any monetary shock under either LP or SP restrictions with any of the three data sets. Similarly, had we imposed all the robust sign restrictions delivered by the models on the data we would not have been able to obtain monetary shocks in any of the 6 cases. These results, together with the previously noted difference in the shapes of the cross correlation functions, suggest that both models have dynamics which are seriously misspecified and are substantially different from those present in the data.

Are the responses of the variables of the system to an identified monetary shock reasonable? In the US, contractionary monetary disturbances identified with LP sign restrictions produce strong negative contemporaneous inflation responses which die out after 4 quarters, a hump shaped decline in output (with a through after 2 quarters) and an initial decline in slope followed by a hump shaped increase. For the Euroland data a contractionary monetary shock produces a positive hump shaped responses in inflation (with the peak response after 4 quarters), a hump shaped decline in output (with a through after 10 quarters) and an immediate decline in the slope. Quantitatively speaking, the persistence of Euroland responses are typically stronger than those found in the US where most of the adjustments take place within 5 quarters of the shocks. With SP sign restrictions, a contractionary monetary shock in the US produces an immediate drop in both output and inflation. However, while inflation returns to its steady state in about 4 quarters, output adjustment takes much longer (about 20 quarters). Since the response of inflation is strong and instantaneous real balances increase to peak after 3 quarters and smoothly decline to the steady state thereafter. In the UK output responses have similar dynamics. However, inflation initially increases, displays an inverted hump shaped response in the medium run with a through after four quarters, and swings around the steady state for several years. Real balances increase but much more smoothly than in the US and the peak response is delayed by approximatively 10 quarters.

To summarize, with both identifications schemes a contractionary monetary policy shock
induces output to decline. The qualitative pattern of output responses is somewhat similar across countries and identifications: although the response is immediate, the through in the responses typically occurs after at least 3 quarters indicating some sluggishness in output adjustments. The qualitative pattern of inflation responses appears to be substantially different across identification schemes: with the SP scheme the sign is, in general, negative, while with the LP scheme the sign of the responses is positive. These responses suggest that monetary policy shocks may be affecting not only the aggregate demand curve of the economy but also its aggregate supply. The extent of this latter effect however appears to depend on the identifying restrictions used. Finally, there appears to be little conformity in inflation responses across countries even within the same identification scheme: in US inflation responses are instantaneous and strong with both schemes, while in the other two countries there is some sluggishness in inflation responses.

4.4 Qualitative Comparison

To qualitatively measure the performance of the two models we examine the dynamic responses to identified monetary shocks for two variables of the system: real balances and labor productivity in the case of the LP scheme and the slope of the term structure and labor productivity in the case of the SP scheme. Although at this stage we restrict attention to these two variables, it should be clear that the use of the dynamic behavior of the other three variables does not create any circularity in the evaluation process since only the sign of their responses, not the dynamic patterns have been used at the identification stage.

There are at least two reasons why a comparison based on these variables may be informative about the properties of the two models. First, we would like to know if the transitory monetary shocks we have identified produce liquidity effects. The generation of liquidity effects has been one of the "tests" used to decide whether a particular identification procedure should be entertained or not in structural VARs (see e.g. Leeper and Gordon (1994)). Recall that, in both models temporary contractionary monetary disturbances produce an increase in short term interest rates, a decline in real balances in the short run and some small inflation effect in the medium run. Since the magnitude of the responses of long term interest rates is negligible, the response of the slope of term structure in both cases is specular to the response of short term interest rates. Since both models have built in liquidity effects we are in the position to verify whether the data constrained with the two types of sign restrictions meet this minimal economic requirement. Failure to produce liquidity effects in the data will be considered an important economic failure of the model.

Second, the behavior of labor productivity in response to shocks has been the focus of attention of several studies in recent years and some authors, e.g. Gali (1999), have used its dynamic response to discriminate between flexible price real business cycle and sticky price demand driven explanation of fluctuations. As we have seen the qualitative behaviour of labor productivity is very similar in the two models: a contractionary monetary policy
shock reduces employment and the reduction is larger than the corresponding reduction in output. Therefore, labor productivity increases. Intuitively, this occurs because when there are (small) costs of adjustments to capital, reductions of the scale production are obtained via adjustments of the more flexible of the two productive factors. Quantitatively, the effect is very short lived in the SP model while it takes about 10 periods for labor productivity to get back to its original steady state in the LP model. Our task here will be to examine whether the data restricted with the two schemes qualitative conforms with the predictions of the two models. A comparison of the persistence of dynamic responses of labor productivity following monetary disturbances may help to discriminate between the two alternative interpretations of cyclical fluctuations.

Figures 7-8 plot the responses of these two variables for each data set (straight lines) together with the responses obtained in the two models (dotted lines). With the LP identification the responses of real balances in the US has the correct sign even though the magnitude of the responses is much more pronounced than in the model. For the Euro data, real balances responses have the wrong sign (no liquidity effect is generated) but the shape of the responses is not too far away from the theoretical ones at least after a few periods. For the SP scheme, a liquidity effect is completely absent in the US while in the UK is very short lived relative to the one produced in the model. In fact, while the slope of the term structure instantaneously decreases in response to a contractionary shock in the data, it becomes positive immediately afterward. There are two possible explanations for this last result: either the movements in short term interest rates are mean reverting or strong expected inflation effects, which drive long term interest rates almost immediately up, are present. Since the output dynamics we have described are inconsistent with the presence of both these features, one must conclude that the liquidity effect generated by monetary shocks in the UK are extremely short lived.

The dynamic responses of labor productivity obtained with the LP identification scheme appear to be at odds with those obtained in the model. Both the sign and the direction of the adjustment of labor productivity responses is incorrect with both data sets: while the model predicts that employment reacts more strongly than output to monetary disturbances, this appears not to be the case in the data. In other words, adjustments in the scale of production following contractionary monetary disturbances do not produce swing in total hours which are larger than those observed in output. This could be due to labor hoarding, sluggishness in employment adjustments, or because other factors (e.g. utilization or inventories) are volatile than output in response to the shocks. Notice, that the smaller response of hours to monetary shocks is not due to the institutional characteristics of various labor markets or to particular rigidities present in the two continents: all data sets deliver the same result. Notice also that while for the US the estimated labor productivity response are somewhat persistent, with Euro land data the effect is extremely short lived and almost negligible. With the SP identification the results are more supportive of the model: for the US the sign of instantaneous labor productivity responses is the same as in the model even though the
dynamic shape is incorrect. For the EUROland data the response of labor productivity is remarkably similar to the one obtained in the model.

To summarize, both models appear to provide an inappropriate representation of the data. First, the sign restrictions imposed by the two models on the joint contemporaneous comovements of inflation, output and the slope (real balances) are inconsistent with the dynamics of at least two data set and this failure generalizes when additional sign restrictions are added. The LP identification can not account for the sign and the shape of the responses of labor productivity in US and Euroland data sets and generate monetary disturbances in the Euro land area which lack liquidity effects. With the SP identification scheme monetary shocks generate instantaneous responses of the slope of the term structure which are at odds with those produced by the model. For the UK there is a liquidity effect but it is very small and lasts only one quarter. Furthermore, the model appears to be at odds with the data as far as expected inflation dynamics are concerned. We will come back on this point in the next section. Finally, and consistent with the results of Gali (1999), the sign restrictions of the SP model produce responses of actual labor productivity which are qualitative more appropriate than those produced using the sign restrictions of the LP model. Furthermore, the persistence of these responses in the data appear to be sufficiently in line with those generated in the model.

4.5 Quantitative comparison

Although the previous subsection has shown that the prediction of the two theories are at odds with the data constrained with restricted derived from the two models - and this would be sufficient to casts doubts about the use of both models for policy and economic analyses - we proceed to undertake a quantitatively evaluation of the two models in order to shed some further light on some of their important features.

There are many ways one could quantitatively measure the discrepancy between the model and the data and there is a number of variables for which the exercises could be performed. We restrict our attention to three statistics, which are simple to calculate and provide economic information which is useful and relevant to potential users of these models. It should be clear, that at this stage our analysis is far from being exhaustive and only attempts to assess the quantitative properties of the models on a number of interesting dimensions.

First, we examine whether the models and the data agree as far as measuring the real effects of monetary policy. For this purpose we calculate first the half life of output in response to monetary disturbances. Chari, Kehoe and McGrattan (2000) have argued that simple SP models of the type we analyze are incapable to generate sufficient persistence in output responses suggesting that price rigidities are insufficient to boost the propagation features of the model. Figure 2 confirms that in this type of models the output effects of monetary shocks die out very quickly. One question of interest is therefore whether the output effects of monetary shocks obtained in the data restricted with SP sign restrictions are
quantitatively similar to those obtained in the model. The statistic we compute gives us some information on this issue. We report results in table 2 where we present the value of the half life obtained in the model and the 68% band of the estimates obtained in the data restricted with the two types of sign restrictions. Bands are computed using a Monte Carlo approach. Intuitively, whenever the bands do include the theoretical value, the data and the model are at odds which each other. The second statistic we report is complementary and measures the percentage of the variance of output 4, 8 and 24 steps ahead due monetary innovations in the two models and in the data. This statistics is typically used by VAR reserachers to assess the contribution of monetary disturbances to business cycle fluctuations: estimates of this percentage in the medium run have varied quite a lot in the literature. We go from 0-20% in studies like Uhlig (1999) or Kim (1999); to 15-30% in studies like Leeper and Gordon (1994) or CEE (1996) to values of 40% or above in studies like Stongin (1995), Faust (1998) and Canova and De Nicolo’(1999). Hence, apart from measuring the relative match of the two models to the data, it is interesting to calculate this percentage to see in which category the data constrained with restrictions imposed by the two models falls. To make the comparison appropriate and since there are five shocks in the VAR and only two shocks in the theory, we augment the two models with three additional shocks (a government expenditure shock, a preference shock and a money demand shock). The persistence and the variance of these additional shocks are reported in table 1. To facilitate the comparison, we report in table 3 68% error bands for these percentages in each of the data set and the value obtained in the models. Once again, if the values produced by the model are outside of this range, the likelihood that the model conforms to the data is minimal (model outcomes are at least one standard deviation away from the data).

Second, we examine whether the link between money and inflation is appropriately accounted for by these two models. To quantitatively measure this link we calculate the cumulative effect of monetary shocks on inflation in the models and in the data and we construct the probability that the value obtained in the two models comes from the data at selected horizons. Table 4 reports the probabilities at horizons 4, 16, ∞: a value outside the range 16-84% (2.5-97.5%) indicates that the cumulative responses of inflation produced by the model is different from the one produced in the data by one (two) standard deviations.

Consistent with the expectations, the half life of the responses in SP restricted data is much longer than the one produced by the SP model. For example, the median half life is 10 quarters in US data and 17 quarters in UK data. In addition, for this last data set, the lower bound of the 68the value of 1 we have in the model. Therefore, we confirm the results of Chari, Kehoe and McGrattan: the persistence of output responses in the SP model is very far away from the one in the data. For the LP scheme, the value of 9 quarters produced by the model is close to the median for the US (the value is 11), while is much too short for the Euro data where where the hump shaped response of output implies very strong persistence in the responses. Overall, and except for the US data constrained with the LP scheme, the persistence of output responses in the data is much larger than the one produced in the
The percentage of output variance accounted for by monetary shocks in US at the 24 step horizon varies between 11 and 43% with the LP scheme and 3 and 34% with the SP scheme. For Euroland output the magnitude of the band is large (from 0.1 to 53%) and for the UK the restrictions imposed by the SP scheme give monetary shocks which account for 9-31% of output variance. In comparison in the two models, and regardless of the exact parametrization of the variance of the different shocks, monetary disturbances account for less than 1% of output variance. In conclusion, both models appear to lack of internal propagation mechanism, and fail to account for the size and the persistence of output responses to monetary shocks. Relatively speaking, the problem is much more severe for the sticky price specification.

The qualitative features of cumulative inflation dynamics in response to monetary shocks are better accounted for by the LP model: for both the US and the Euro land the sign of the cumulative responses is correct. This is not the case for the SP model, where the sign of cumulative inflation responses are consistently on the wrong side for both data sets. This qualitative feature is confirmed by the probabilistic results presented in the table. For example, the responses of the LP model in the long run are very near the median value of the estimated band and at no horizon do the cumulative responses of the model fall outside the 95% range for both data set. This is not the case for the SP: only in the long run the value obtained in the model falls within the 95% range for the US and only in the very long run for the UK. Furthermore, with both data set the probability to draw a cumulative value in the data of lower magnitude than the one produced by the model at short horizons is negligible. Hence, the uncertainty present in estimating cumulative inflation responses in the data is not sufficient to reconcile both sign and size difference between the model and the data.

5 Other implications of the models

The analysis far we has concentrated attention on monetary disturbances because, as mentioned, the major difference between the two models is in the way monetary disturbances affect the real side of the economy. However, our approach has not been tailored only to the identification of monetary disturbances and can be used also to extract any types of disturbances for which there some relevant theoretical counterpart. Therefore the approach provides alternative and complementary channels through which the comparison between the models and the data can be undertaken. To illustrate this point we examine here the relative performance of the two models when the sign restrictions derived in response to technology disturbances are used to identify these shocks in the data.

As it was the case with monetary shocks, there are several robust restrictions produced by the two models in response to technology disturbances. For the sake of symmetry, we focus
once again on the sign restrictions imposed on output, inflation and the slope in the case of
the LP scheme and output, inflation and real balances in the case of the SP scheme and choose
the rotation of reduced form shocks which simultaneously identify the monetary shocks we
have discussed in the previous subsections and the technology disturbances we analyze here.
Figure 9 reports the pairwise cross correlation functions of these three variables in response
to technological disturbances in the models, figures 10 and 11 the pairwise cross correlation
functions obtained in the data using the contemporaneous sign restrictions presented in
figure 9 and figures 12 and 13 the impulse responses of the five variables of the systems
to the identified shocks. Note that also in this case, imposing the whole vector of sign
restrictions results in outright rejections of both models.

Few features of the figures deserve some discussion. First, no technology shock is identi-
fiable in the UK with the SP scheme. Second, the responses of output, inflation, the slope
in the all data sets are remarkably similar with both identification schemes. Since, it can be
seen in figures 1 and 2 that the two models have very similar implications for these three vari-
ables in response to technological disturbances, it is comforting to observe that the responses
in the data display similar features. Clear differences however emerge in the responses of
real balances and labor productivity: real balances always decline after a positive technology
with the SP scheme while with the LP scheme they decrease in the US and increase in the
UK and Euroland. The negative response of real balances is counterfactual, it is at odds
with the responses produced in the model and highlights a substantial mismatch between
the theory and the data footnotelIt is however possible that the odd behavior of real balances
in the data is due to measurement errors or to the fact that M1 is a poor counterpart of
cash balances in the model. Labor productivity responses also have signs which depend
on the identification scheme and the country. For example, in the US responses are positive
for the LP scheme and negative for the SP scheme, while in the Euroland the opposite is
ture. Note in theory this was not the case - labor productivity responses were always posi-
tive, as hours increase by less than output. Hence, at least with one data set both schemes
have hard time to account for the behavior of labor productivity. Moreover, while the SP
model was much more adequate to explain the behavior of labor productivity in response
to monetary disturbances its performance is as mixed as the one of the LP model in the case
of technological disturbances.

In both models technology shocks are the dominant source of real fluctuations. One
may be curious as to if this is the same also in the data once it is restricted with the sign
restrictions derived by the two schemes. The answer is positive. The value obtained in the
model at the 24 step horizon (57%) is within one standard error band of the percentage
found in all data sets with all identification schemes. Therefore, both models provides an
accurate representation of the propagation mechanism of technology disturbances to the real
side of the economy.
6 Improving the match between the model and the data

Both models display important shortcomings in explaining the dynamic response of the variables of the system in the data. One may be therefore interested in going back to the drawing board and examining whether there are changes in the structure of the models which may improve the quality of the approximation. In this section we concentrate attention on one particular extension of the setup we have used and study whether the inclusion of this feature improves the qualitative match of the model to the data.

We have seen that the behavior of labor productivity in response to monetary disturbances when the data is restricted with the LP scheme is inconsistent with the one produced by the model. We have also argued that this is due to the fact that the response of hours tends to be larger than the response of output in the data while this is not the case in the model. One question worth investigating is whether it is possible to introduce some new feature, which maintains the sign restrictions we have previously derived, and changes the sign of the dynamics of labor productivity in response to monetary shocks. The modification we consider is the introduction of factor services utilization.

Capacity utilization is a natural candidate to consider since this feature has been used in recent papers to make the dynamics of DGSE models in response to shocks more similar to the one found in the data (Burnside and Eichenbaum (1996), Neiss and Pappa (2001)). We add capacity utilization to the LP model in a very simple way: we assume that the production function depends on labor, the amount of capital available and its utilization. Given a stock of capital, varying capacity utilization (through hoarding or overtime) allows output to vary in response to shocks. If capacity utilization varies enough, it may be possible to reduce the impact of monetary shocks on employment and therefore make labor productivity fall in response to the shocks. We assume that varying capacity utilizations has costs: the more the capital is utilized, the faster it depreciates. We summarize the alterations produced by this features to the model in appendix B. Is this modification sufficient to bring the model more in line with the data? We first confirmed that a model with capacity utilization produced the sign restrictions we have used to identify monetary shocks in the data for a range of values of the parameter which controls the depreciation cost due to capital utilization. Second, we searched within this range for a the value of the parameter which gave us a negative response of labor productivity. We found none. In figures 7 we show what happens to labor productivity when this parameter is fixed to the standard value of 1.56: labor productivity responses are more sluggish, as intuition would suggest, but by itself capacity utilization it is not sufficient to revert the relative magnitude of the changes in output and hours in response to monetary shocks.

Next, we consider splitting employment into the two margins: hours worked (intensive) and number of people working (extensive). We assume that there is sluggishness in employ-
ment decision by requiring the supply of bodies (the proportion of total population seeking employment) to be decided one period in advance. Hence, firms may alter effort in response to the shocks but not employment, at least instantaneously. Hence, we treat labor as we have treated capital in the previous exercise, allowing some overtime/hoarding when the economy is perturbed by shocks. Ee summarize the alterations of the model produced by this feature in appendix B. Also in this case, the joint dynamics of output, inflation and the slope are qualitatively unaltered by allowing the employment decisions to be taken on the two margins. (To be continued)

7 Policy Analyses

Based on the results we have obtained so far, both models appear to be unsuitable to undertake policy analyses: the dynamics they produce are at times both qualitatively and quantitatively inconsistent with those produced by the data once restrictions provided by the models are imposed. Despite this general failure one may be nevertheless interested in knowing, in the spirit of King and Wolman (1999), what these models have to say about the welfare properties of different monetary policy rules. To do this we examine how the utility of the representative consumers is displaced on the transition path by unexpected contractionary monetary shock under three different rules in the two models: a taylor rule, a partial accommodative rule linking interest rates to real balances and an inflation targetting rule. Although no claims of optimality can be made, our exercise can highlight the relative ranking of these three policies. In each case we calculate the compensating variation of consumption needed to agents to remain on average on their steady state path. In other words we compute:

$$0 = \sum_t \beta^t \frac{(c_t - \bar{c}_{LP})^\mu (1 - n_t)^{1-\mu})^{\gamma}}{\gamma}$$

and

$$\sum_t \beta^t \frac{(c_t - \bar{c}_{SP})^\mu (1 - n_t)^{1-\mu})^{\gamma}}{\gamma} + \frac{\theta}{1 - \nu} \frac{(M_{t+1})^{1-\nu}}{p_t}$$

where $\beta = 1.04^{0.25}$, $\gamma = -2.0$, $\mu = 5.0$, $\theta = 1.0$, $\nu = 7.0$. A policy rule is preferable if it requires smaller (in absolute value) compensating variations in consumption. For the LP model the smaller compensating variation is obtained with the partial accommodative rule (-0.0050) followed by the taylor rule (-0.0052) and the inflation targetting rule (-0.017). For the SP model the partial accommodative rule is the once again preferred in terms of compensating variation (0.0028) while for the other two rules the order is reversed (0.0047 for the inflation targetting rule, 0.0054 for the taylor rule). The intuition for the preferability of partial accommodative rules is clear: consumption and employment are less volatile with this rule because real balances show a much smoother behavior in response to shocks than either
output or inflation. The difference in the ranking between taylor and inflation rules in the two models can be explained by the behavior of output responses which is much smoother in the LP model. In conclusion, both frameworks of analysis indicate that partialacommodative rules are to be preferrable for welfare stabilization. Quantitatively the gains are small in the LP model but sizable and significant in the SP model.

8 Conclusions

This paper provides an alternative methodology to compare DSGE models to the data. We argued that the procedure is simple, easily reproducible and does not face some of the problems encountered with existing approaches; it allows to evaluate the model in several dimensions in response to different shocks and to respecify the theoretical construction at any level of the evaluation when the discrepancy between the model and the data is large; it is designed to provide both qualitative and quantitative comparisons and is suited to answers interesting economic questions which a researcher may want to consider.

The procedure takes seriously the objection that DSGE models are, at best, approximations to the DGP of the actual data and even as approximations they may be appropriate only in a few dimensions. We also are sympathetic to the idea that economic validation (as opposed to statistical validation) is what matters for users of DSGE models and we propose an evaluation approach which takes these concerns into account. Finally, the procedure incorporates in the analysis the objection that too little sensitivity analysis is typically performed on calibrated DSGE models - and therefore that the outcomes of the model may depend on particular and arbitrary parameter choices. In fact, the first step of our approach searches for robust implications of the model, implications which hold regardless of the exact parametrization, of particular functional choices or of additional features that a researcher may want to include in the specification or not. Once robust implications are discovered, we used them as identification devices to construct structural shocks in VAR models. That is, we transform a reduced form VAR into a structural model by imposing a minimal set of robust restrictions derived from a DSGE model. We then use responses to identified shocks in the data and in the model to provide a metric to compare the outcomes both qualitatively and quantitatively. Based on the results obtained, one can then proceed to respecify the model when the discrepancy is large, answer interesting economic questions or conduct policy analyses of some sort.

We use this machinery to evaluate two, by now standard, DSGE models of money - a limited participation economy and a sticky price monopolistic competitive economy - against the data. We show that the models are at odds with the data in several ways. First, for some data sets the robust restrictions derived by the model do not have counterparts in the data. In other words, some fundamental implications of the model are qualitatively inconsistent with the data. Second, we have shown that qualitatively the dynamics of both
models are different than those in the data. For example, the limited participation model has hard time to replicate the sign of the responses of labor productivity following a monetary policy shock and the sticky price economy generate dynamics for inflation and the slope of the term structure which are inconsistent with the presence of liquidity effects. Third, monetary disturbances in the sticky price model generate output responses whose persistence is quantitatively different from those obtained in the data, once SP restrictions are used. Moreover, both models fails to account for the size of the real effects due to monetary policy shocks in the data. Also, the limited participation model quantitatively capture the effects of policy shocks on inflation better than the sticky price model. Fourth, while the two models generate dynamics in response to technology disturbances which are qualitatively and also quantitatively similar, the data constrained with the models’ restriction produce responses which are fragile in the sense that they depend on the data set used and to some extent on the set of restrictions we impose. Finally, the two models have different implications regarding the relative ranking of preferred policy rules even though they agree that a partial accommodative one is preferable on pure welfare grounds.

Overall, our exercise sheds new and important light on the nature of these models and on their match with the data. We believe this information can be useful to theorists engaged in respecifying some of their features and to policymakers who should decide which theoretical economy should used in taking policy decisions.
Appendix A

In this appendix we describe how to implement the sign restrictions derived from the model in the data. The approach uses the following two results:

Result 1: Let $P$ be the matrix of eigenvectors and $D$ the matrix of eigenvalues such that $\Sigma = PDP'$. Then $P = \Pi_{m,n} Q_{m,n}(\theta)$ where $Q_{m,n}(\theta)$ are rotation matrices of the form:

$$Q_{m,n}(\theta) = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cos(\theta) & \cdots & -\sin(\theta) & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \sin(\theta) & \cdots & \cos(\theta) & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

where $0 < \theta \leq \frac{\pi}{2}$ and the subscript $(m, n)$ indicates that rows $m$ and $n$ are rotated by the angle $\theta$.

Result 2: If $P$ be the matrix of eigenvectors and $D$ the matrix of eigenvalues of a matrix $\Sigma$ so that $\Sigma = VV'$ where $V = PD^{0.5}$ then also $\tilde{V} = PD^{0.5}Q(m, n)$ satisfies $\Sigma = \tilde{V}\tilde{V}'$.

In our case $\Sigma$ is the covariance matrix of reduced form VAR disturbances and the second result holds because $Q(m, n)$ are orthonormal. The algorithm works as follows:

- construct matrices $P$ and $D$, check if that decomposition one or more the disturbances of the system produce the required sign restrictions on the endogenous variable of the VAR. If yes, stop search.
- If not construct a matrix $P(m, n)(\theta) = P \ast Q(m, n)(\theta)$ for some $m, n$ and some $\theta$ and check if under the new decomposition one or more shocks satisfy the required sign restrictions. If yes, stop, if the restrictions are not satisfied try another pair $m, n$ for a given $\theta$ or another $\theta$ for any pair $m, n$. In a system with $s$ variables there are $(s(s-1)/2)$ pariwise rotations $(s(s-1)/4)$ combinations of pairwise rotations, etc. for any given $\theta$. So for example in a system with $s=5$ variables, there are 10 pairwise rotations. Note that, each pariwise rotation "decouple" the matrix of eigenvector in a particular direction. To select $\theta$ we grid the interval $[0, \frac{\pi}{2}]$ into $M$ points, $M$ typically large. In practive we construct $M \ast (s(s - 1)/2)$ orthogonal decompositions of $\Sigma$, and store all those which satisfy the required sign restrictions. It is typically the case that there is a range of $\theta$ which satisfy the sign restrictions for a given $m, n$. Since the impulse responses within this range are similar, we consider this range a point. If there are different ranges which satisfy the restrictions or different combinations $m, n$ for one $\theta$ we choose the range or the pair which maximizes the number of interpretable shocks. Canova and Denicolo' (1999) provide further steps to narrow down the class of admissible decompositions in case this was not sufficient to select one candidate. When there is more than one shock which satisfies the restriction we choose the one which produces responses.
with the highest conformity with the model (both in terms of shape and size of the responses). In general, when more than one shock produces the same sign comovements in the data, it is advisable to add more sign restrictions or more variables to attempt to disentangle them. The interested reader is invited to consult of Canova and De Nicolo' (1999) for the details.

### Appendix B

In this appendix we describe the modifications needed to account for variable factor utilization in the LP model. Consider first variable capital utilization. In that case we write the production function as $y_t = v_t(u_t k_t)^{\alpha} n_t^{1-\alpha}$ where $u_t$ is capital utilization. The law of motion of capital is $x_t = k_{t+1} - (1 - \delta u^\phi)k_t + x_t$, where the parameter $\phi$ controls the costs of overutilization in terms of depreciation (it is the elasticity of depreviation to utilization). Note that for $\phi$ large capacity utilization is constant.

Next, consider the case of variable employment utilization. The production function in this case is $y_t = v_t(k_t)^{\alpha} (e_t H_t)^{1-\alpha}$ where $e_t$ is effort and $H_t$ employment. We modify the utility function of the representative agent to be

$$\max \left\{ c_t, e_t, n_t, k_t + 1, I_t \right\} \quad E_0 \sum_t \beta^t \frac{(e_t^\mu (1 - e_t)^{1-\mu} H_t^{\theta})^\gamma}{\gamma}$$

(15)

where $\theta$ a parameter. Note that as $\theta \to 0$ the specification collapses to the previous one. We make the assumption that employment is predetermined (is chosen one period in advance) so that only effort instantaneously respond to shocks. The constraints faced by the consumers are now:

$$c_t p_t \leq M_t - I_t + W_t H_{t-1} e_t$$

(16)

$$M_{t+1} = D_t + F_t + r_t p_t k_t + M_t - I_t + W_t H_{t-1} e_t - c_t p_t - x_t p_t$$

(17)

$$x_t = k_{t+1} - (1 - \delta)k_t + \frac{a}{2} \left( \frac{k_{t+1}}{k_t} - 1 \right)^2 k_t$$

(18)
References


[22] Ireland, P., 1999, A method for taking Models to the Data, Boston College, manuscript.


[27] Geweke, J., 1999, "Computational Experiment and Reality", University of Iowa, manuscript.


[29] Uhlig, H., 1999,
<table>
<thead>
<tr>
<th>Table 1</th>
<th>Parameter Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common Parameters</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>( N_s )</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Parameters Limited Participation Economy</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>m/c</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Parameters Sticky Prices Economy</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \epsilon )</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td><strong>Standard deviation of various shocks</strong></td>
<td>technology</td>
</tr>
<tr>
<td></td>
<td>US</td>
</tr>
<tr>
<td></td>
<td>UK</td>
</tr>
<tr>
<td></td>
<td>Euro</td>
</tr>
</tbody>
</table>

Notes: Values for Euro land are taken from Smets and Wouters (2001), the other are obtained from several sources.
### Table 2

**Half life of output responses**

<table>
<thead>
<tr>
<th>Model</th>
<th>Value</th>
<th>US</th>
<th>UK</th>
<th>Euro</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP</td>
<td>9</td>
<td>[1,12,40]</td>
<td>[40,40,40]</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>1</td>
<td>[2,10,40]</td>
<td>[6,17,40]</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** In each both the first number is the 16% percentile of the distribution, the middle the 50% and the third the 84% percentile. Half life is measured in quarters.

### Table 3

**Explanatory power of Monetary shocks for output**

<table>
<thead>
<tr>
<th>Step</th>
<th>Model</th>
<th>US</th>
<th>UK</th>
<th>Euro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Limited Participation economy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>[0.41, 0.45]</td>
<td>[0.001, 0.13]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>[0.30, 0.045]</td>
<td>[0.001, 0.34]</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.01</td>
<td>[0.11, 0.44]</td>
<td>[0.006, 0.52]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sticky Price economy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>[0.05, 0.34]</td>
<td>[0.20, 0.36]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>[0.02, 0.34]</td>
<td>[0.14, 0.35]</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>0.01</td>
<td>[0.03, 0.33]</td>
<td>[0.09, 0.31]</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** In each both the first number is the 16% percentile of the distribution and the second the 84% percentile of the distribution.

### Table 4

**Probability for cumulative responses of inflation**

<table>
<thead>
<tr>
<th>Step</th>
<th>US</th>
<th>UK</th>
<th>Euro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Limited Participation economy</td>
</tr>
<tr>
<td>4</td>
<td>0.98</td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td>16</td>
<td>0.70</td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.54</td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sticky Price economy</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.95</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.70</td>
<td>0.90</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Each cell reports the probability that cumulative responses of inflation in the model are less than in the data.
Figure 1:
Figure 2:
Cross Correlations, monetary shock

Model LP

Figure 3:
Cross Correlations, monetary shock

Model SP

Figure 4:
Cross Correlations: Monetary shocks

Sample 80:1-98:4, Identification LP1

Figure 5:
Figure 6:
Cross Correlations, Monetary Shocks

Sample 80:1-98:4, Identification SP1

Figure 7:
Responses to Monetary Shocks
Sample 80:1-98:4, Identification SP1

Figure 8:
Cross Correlations, technology shock

Figure 9:
Cross Correlations, Technology Shocks

Sample 80:1-98:4, Identification SP1

Figure 10:
Responses to Technology Shocks

Sample 80:1-98:4, Identification SP1

Figure 11:
Cross Correlations: Technology shocks

Sample 80:1-98:4, Identification LP1

Figure 12:
Responses to Technology Shocks

Sample 80:1-98:4, Identification LP1

Figure 13: