DISCUSSION PAPER SERIES

No. 3442

VALIDATING MONETARY DSGE MODELS THROUGH VARS

Fabio Canova

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP3442.asp**

ISSN 0265-8003

VALIDATING MONETARY DSGE MODELS THROUGH VARS

Fabio Canova, Universitat Pompeu Fabra, Barcelona and CEPR

Discussion Paper No. 3442
July 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3442

July 2002

ABSTRACT

Validating Monetary DSGE Models through VARs*

Robust sign restrictions derived from calibrated DSGE models are used to identify structual shocks in the actual data. The dynamic behaviour of selected variables in response to these shocks is employed to measure, both qualitatively and quantitatively, the economic discrepancy between the model and the data. We design an algorithm that allows increasingly demanding diagnostics on the model, room for respecification at each stage of the process and comparison across models. We show that neither a limited participation model, nor a sticky price monopolistic-competitive model, fully accounts for the dynamics of a small set of macro variables. Furthermore simple alterations of the former fail to improve the match with the data, even in qualitative sense.

JEL Classification: E00 and E50
Keywords: dynamic general equilibrium models, model evaluation, sign restrictions and VARs

Fabio Canova
Department of Economics
Universitat Pompeu Fabra
Ramon Trias Fargas, 25-27
08005 Barcelona
SPAIN
Tel: (34 93) 542 2659
Fax: (34 93) 542 1746
Email: canova@upf.es

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=114830

* This Paper grew out of the Panel Discussion of the workshop ‘SDGE Models and their use in monetary policy’, held at the European Central Bank. I would like to thank J Valles, M Binder, C Sims and the participants of the TSM conference in Toulouse, the workshop ‘VAR for monetary policy’ at the Riksbank, the 4th workshop ‘New Approaches to the study of economic fluctuations’, Hydra and of seminars at Federal Reserve Bank of Richmond, Federal Reserve Bank of Atlanta, Bank of England, University of Maryland, LSE and LBS for comments and suggestions. This Paper is produced as part of a CEPR Research Network on New Approaches to the Study of Economic Fluctuations, funded by the European Commission under the Training and Mobility of Researchers Programme (Contract No ERBFMRX-CT98-0213).

Submitted 12 June 2002

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models and structural VARs have acquired a permanent status in the toolkit of macroeconomists in the last 15 years. Although authors disagree on the exact microfundations to be used and on the frictions needed to make DSGE models 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 has been refined over the years and many features have been added to the benchmark specifications 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, VARs analyses eschew most of the detailed microstructure underlying DSGE models and attempt to draw inference on the most likely effect of structural shocks by imposing only a minimal set of theoretical constraints. Validation in these models is typically undertaken using a variety of statistical and economic procedures, which at times results 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 a class of models is more appropriate than another to explain some phenomena or whether the introduction of an additional feature improves the match of a model to the data. Policymakers, on the other hand, need to make sure that a model approximates sufficiently well the data if they want to use it for policy purposes. To lend support to a specification one wants to entertain or to select a particular candidate for policy exercises, validation techniques which are simple, reproducible, effective in comparing the economic discrepancy between the model and the data and informative on the reasons when differences emerge are clearly needed.

One reason for why validation methods for DSGE models are still at an infant stage, and informal measures are typically preferred, has been an earlier philosophical dispute about the intrinsic nature of an economic model. For example, it was claimed that existing econometric tools are appropriate to measure the fit when the model represents the true DGP of the actual data (i.e. the discrepancy between the model and the data is a random noise with mean zero and 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 are useless because "we reject a model known to be false". In this situation "computational experiments", as defined by e.g. Kydland and Prescott (1996), and its underlying calibration exercise, substitute the standard set of econometric techniques popular in the 70s and the 80s. Two aspects of Kydland and Prescott's criticisms are well taken: the futility of applying standard methods when the model is believed to be a false description of the data; and the need to design criteria which put special emphasis on economic discrepancy as opposed to statistical one. However, the insistence on rejecting econometric tools per-se has led to a partial stall in the development of useful validation techniques. The literature has incorporated to some extent the ideas put forward by followers of computational experiments within well established econometric approaches (see e.g. Watson (1993), Canova (1994), DeJong, Ingram and Whitemann (1996) and Geweke (1999) among others), but several problems still remain.

The task of this paper is to provide a new methodology to validate DSGE models that integrates "calibration" and VARs approaches and attempts to overcome some of the problems present in the existing literature. The starting point is a class of DSGE models which have an approximate state space representation once (log-)linearized around their steady states. We take seriously the objection that such models are at best approximations to the DGP of the actual data and that the approximation may be valid only in some dimensions. We also take seriously the idea that economic validation is what matters for users of DSGE models and we design a 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 may depend on particular and possibly arbitrary parameter choices. For this reason, our approach searches for robust implications of the model, implications which hold regardless of the exact parametrization, of the particular functional choices or of other standard features that a researcher includes in the specification. Once robust implications are discovered, they are used as devices to identify structural shocks in VAR models. That is, we "calibrate" the data to the model by imposing on the VAR 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 an economic metric to qualitatively and quantitatively compare dynamic outcomes. 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 if this is the scope of the exercise.

There are several advantages of our proposed approach: it is simple and easily reproducible; it integrates the ideas present in computational experiments and VAR approaches in a framework where both become useful; it focuses on implications which are robust to the exact calibration of the DSGE model and to assumptions about certain functional forms; finally, and contrary to existing procedures which are designed to recover only one particular type of shock, it can be used to compare models using the dynamic response to several types of shocks.

We apply this methodology to the problem of evaluating two, by now standard, DSGE 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 and 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, only few isolated studies examine the performance of these models in relation to each other or in comparison with other structures for a common set of problems (see e.g. Tack (1996), Christiano, Eichenbaum and Evans (1997), Schorfheide (2001)). 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 both models are at odds with the data in several ways. First, for some countries the robust restrictions that these models 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 certain variables 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 generates a dynamic shape for the slope of the term structure which are not fully consistent with those present in the data. Third, quantitatively speaking, the match between the models and the data is poor. For example, the magnitude of output responses to monetary disturbances in both models are small relative those found in the data. In addition, in response to policy shocks, the sticky price model produces real effects which lack the persistence found in the data (see Chari, Kehoe and McGrattan (2000)) and it is unable to match inflation dynamics that result from these shocks. Fourth, dynamics in response to certain shocks are fragile, in teh sense that they depend on the data set used. This is the case, for example, for both models when Technology shocks are considered.

We show that two simple alterations of the basic the limited participation structure are unable to bring the responses of labor productivity in line with those observed in the data, suggesting that more drastic modifications are needed to make the theory consistent with the data in this dimension. Finally we demonstrate that, although the mechanics of the propagation of shocks are different, the two models have similar implications regarding the choice of particular policy rules: consumer's welfare is highest when nominal interest rate systematically responds to real balances.

The rest of the paper of the paper is organized as follows: the next section briefly reviews the literature, describing existing approaches to evaluate DSGE models and pointing out their weaknesses. Section 3 describes in details the alternative methodology. Section 4 presents the results of the evaluation of two monetary DSGE models. 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 emerged in the literature. This section briefly summarizes the current state of the art, compares various approaches and highlight unsolved problems.

The oldest approach builds on work by Sargent (1978) and Altug (1989) and takes a 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 with nuisance features, i.e. adding dynamics (see Diebold, Ohanian and Berkowitz (1998)), measurement errors (Ireland (1999)) or shocks (e.g. Ingram, Kocherlakota and Savin (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 specification is presumed to correctly represent 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 examine the quality of the model. Statistical testing has been recognized to provide little information about economic relationship. Therefore, formal statistical methods are at times employed in conjunction with informal ones, and the information concerning moments, cross correlation functions, and impulse responses is used to complement the results of statistical tests (see e.g. Kim (2000)).

There are several problems connected with this approach. First, the way the probabilistic structure of the model is completed is arbitrary and leaves room for non-comparability across studies. Second, both when the probabilistic structure of the model is completed and when it is not, computational problems typically emerge. Problems of this type occur when the likelihood function has multiple peaks or flat areas, or when information about some parameters is "weak". Third, parameter estimates often turn out to be unreasonable from an economic point of view and lie on the boundary of the admissible parameter space. These latter two problems taken together indicate that misspecification is still present and that statistical tests, which take the model to be correct under the null, are 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 final goal of the exercise.

Recently, and to avoid to report unreasonable parameter estimates, several authors have used a mixed approach, where some parameters are a-priori selected and other are estimated. This standard quick fix makes computations easier, since the restricted likelihood function for the restricted model is typically better behaved, but may jeopardize the validation exercise, since the distribution of estimated parameters non-trivially depends on the values for the parameters which are fixed a-priori. Gregory and Smith (1991) showed, in the context of a simple equity premium model, that the distribution of estimates are skewed and wrongly centered when parameters which are fixed a-priori inconsistently estimate true ones. More recently, Leeper and Zha (2000) demonstrate that inference concerning the magnitude of the inflation coefficient in a Taylor rule is not independent of the values assumed for parameters in other parts of the model. Hence, impulse responses, cross correlations or moments computed from models estimated this way are likely to misrepresent the dynamics obtained when all parameters are jointly estimated.

A second strand of literature takes the view that a DSGE model is false even when the dynamic specification is enriched in various ways. That is, a model is a stylized description of reality 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 full information maximum likelihood is unlikely to be successful because of the misspecification present, and even when this is possible, parameter estimates are unlikely to be meaningful, making statistical validation uninformative - the model will be rejected no matter what.

According to the most extreme version of this view, traditional econometrics should be dispensed of and economic questions should be answered by means of computational experiments (see e.g. Kydland and Prescott (1996)). Validation is implicit in the design of the experiment since the model fits certain 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 other features of the data and the theory's credibility is enhanced if a larger number of stylized facts are matched.

Despite statements suggesting the contrary, computational experiments are designed and performed using a well defined statistical criteria, and validation is based on a loss function which weighs only first moments (see e.g. Canova (1994)). Several criticisms have been raised in the literature regarding both the theoretical foundations and the outcomes of the computational experiments. First of all, the process of calibrating a model to the data typically leaves a number of parameters unspecified and therefore induces arbitrariness in comparing across studies. Second, failure to provide measures of uncertainty to the outcomes of the experiments is seen as detrimental to the establishment of secure foundations to the approach (see Hansen and Heckmann (1996)). Recently, however, Broze, Dridi and Renault (1999) have provided choice theoretical foundations to computational experiments and showed when it is optimal to informally evaluate a model.

A less extreme point of view, maintains that a DSGE model is a false representation of reality, but indicates that it may provide a reasonable description for certain aspects of the data. Hence, instead of estimating/evaluating a model as if it provided a description of the likelihood of the events observed in the data, this approach suggests that one should be modestly concerned with the more limited task of evaluating the model on those specific aspects 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, uses the R^2 of the regression of the data on the model as a metric to measure discrepancy and does not allow variability in the parameters. Christiano and Eichenbaum (1992) and others use the *sampling variability of the actual data* to provide GMM style estimate of the parameters of interest and goodness of fit tests Other authors have used *sampling variability of the model* to construct simulated method of moment (SMM) estimates of the unknown parameters and confidence bands/probabilistic statements for the goodness of the model (see e.g. Gregory and Smith (1993) or Canova and Marrinan (1996)). More recently, approaches which take a symmetric view about the uncertainty surrounding the actual and model's data emerged (see e.g. DeJong, Ingram, Whitemann (1996)) and Bayesian methods have been developed to validate models in a situation where even the few moments used to estimate/ evaluate may be misspecified (see e.g. Geweke (1999) and Schorfheide (2001)).

Several problems have been highlighted with this class of approaches. For example, Pagan (1994) and Geweke (1999) argued that there are logical inconsistencies in viewing the model as a representation of only some dimensions of the data. These inconsistencies can be solved if the DSGE is taken to be a representation for the moments of the data (not of the data itself). In addition, validation methods are built on the assumption that the model truly characterizes some aspects of the data, an assumption which is probably as heroic as the one that the model is the correct DGP for the entire data set. Furthermore, the problems faced by mixed approaches are more acute in these setups because a researcher has a degree of arbitrariness in selecting the features of the data which are used as a benchmark against which the model is evaluated (see e.g. Tauchen and Gallant (1996)). Moreover, most of the approaches rely on measures of fit which are, at best, only weakly informative for the 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 complicated. Computational constraints become binding and existing methods turn out to be impractical or hard to deal with when parameter space is large (say, greater than 5) or number of time series under consideration is moderately large (greater than 4-5).

3 The procedure

The approach we propose attempts to avoid the majority of these problems. We start from the presumption that DSGE models are too stylized to be taken seriously, even as an approximation to part of the DGP of the actual data. Given that the model is potentially misspecified, our task is first to find a set of implications which are robust to the exact parametrization and to arbitrary choices for the functional form of the primitives of the model, and second to use them to restrict the actual data (see also Canova, Finn and Pagan (1994)). Once the data has been "calibrated" to look like the model in some dimensions, we attempt to qualitatively and quantitatively examine the mismatch between the model predictions and the data in other dimensions. The discrepancy is measured using both economic and statistical criteria: the propagation mechanism of shocks is used to provide an economic metric to assess the difference between the model and the data and Monte Carlo methods are used to probabilistically establish the significance of these differences. If the mismatch is not too large, policy analyses can be undertaken and answers to interesting questions provided. Otherwise, one can go back to the drawing board, respecify the model and repeat the exercise.

Our approach shares 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 DSGE model. This is an important step which sets our approach apart since conventional exclusion restrictions used in VAR analyses are at times inconsistent with the dynamics of DSGE models (see Canova and Pina (2000)). Moreover, we consider only restrictions which are robust from the point of view of the model. As in calibration exercises, we use some observations to make the data and the model look alike, but we reverse the existing practice by using restrictions derived from the model to " calibrate" the data. Contrary to the literature, we measure economic outcomes both qualitatively and quantitatively. Also, contrary to most of the analyses, we examine structural dynamics in response to shocks instead of comparing reduced form moments of the model and the data.

Our approach is flexible and permits the identification of every shock which produces robust dynamics in the model. This distinguishes it from existing procedures which typically allow, at most, the evaluation of a model in response to only one shock. Since we can simultaneously analyze the implication for several time series of different structural disturbances (e.g. monetary, technology, fiscal, etc.) our approach allows a more thorough evaluation of the properties of the theoretical economy. Finally, the method is straightforward, computationally simple and easily reproducible.

The procedure involves six specific steps:

1. Find robust implications of a class of models.

2. Use these restrictions to identify shocks in the data. If restrictions do not hold, stop evaluation.

3. If restrictions are satisfied, evaluate the performance **qualitatively** by comparing responses of selected endogenous variables to shocks in the model and in the data.

4. Cross validate **qualitatively** different classes of models, if needed.

5. If discrepancy in 3.-4. is not too large, and policy analyses are the goal of the exercise, continue validation process **quantitatively**.

6. Respecify the model if performances in either 2. or 3.-4.-5. is unsatisfactory or undertake policy analyses, welfare calculations, etc. as needed.

The first step of our procedure is explicitly designed to cope with the inherent arbitrariness of calibration procedures. We define an implication to be robust if it holds independent of parametrization and of 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, we call this a robust implication. Robustness is not generic since many features of a dynamic model are sensitive to what is included or excluded from it. What we are looking for here is a set of implications which is representative of the class of models we want to evaluate. As we will see later on, robust implications typically take the form of shape or sign restrictions. At times, there may be magnitude restrictions which are robust. However, given that models are misspecified, magnitude restrictions are unlikely to be satisfied in the data. Finally, while both unconditional and conditional comovements can be used, we find it more informative to examine comovements conditional on particular shocks, since empirical analyses are typically concerned with the way endogenous variables react to structural shocks.

The second step involves making the model and the data look alike in some aspects. In particular, we would like to identify disturbances in the data which generate the same dynamic responses for certain variables as shocks in the model. Clearly, it is possible that restrictions derived from the model do not hold in the data. In that case, one would either repeat the exercise imposing alternative but similarly robust restrictions, or, if all robust implications are examined and no shock with the required properties yet found, stop the evaluation process and go back to the drawing board.

The third step is similar to that employed in many evaluation procedures. 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: robust implications are used to identify shocks; the sign and the shape of the dynamic response of other variables to shocks are used to check the quality of the model. 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 shape of the response of capital or employment to this shock is qualitatively similar in the model and in the data. We stress that at this stage the evaluation is qualitative.

At times a researcher is not necessarily interested in the absolute performance of a model but instead, may be concerned with the relative likelihood of models which differ in some relevant features, for example microfundations or economic frictions. These features can be nested (a model with capacity utilization or without) or non-nested. Once the first three steps of our evaluation procedure have been undertaken and none of the models immediately set aside, it is then possible to qualitatively compare models. For example, one may ask whether the restrictions imposed by model X are more appropriate than those of model Y in making output dynamics in response to monetary shocks similar to those in the data, or whether the addition of a feature Z to the model enhance the match. Once again, we keep the comparison qualitative: several models can be discarded using qualitative devices such as sign and shape features of the responses.

In many cases the final scope of the analysis is to give answers to certain quantitative questions, to undertake some policy exercises or perform welfare calculations. For instance, one may want to measure the size of the real effects of a nominal shock; assess whether welfare losses due to distortions and frictions are larger in a model or another (as e.g Smets and Wouters (2001) or know whether Taylor-type policy rules induce movements in macrovariables in response to shocks which are different than those produced by inflation rules. In all these cases Monte Carlo methods can be used to assess the statistical significance of the results, to construct probabilities of interesting events (as e.g. Canova (1994)) and to evaluate the magnitude of welfare losses.

4 Evaluating two monetary models

To illustrate how the procedure works in a concrete example, we examine how good two standard monetary DSGE models are in explaining the dynamics observed in the data. The first model is a version of the Limited Participation economy studied, among others, by Christiano, Eichenbaum and Evans (CEE) (1997) and Canova and Pina (2000). The second is a version of a monopolistic competitive model with sticky prices examined, among others by 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 will be distorted if this feature is excluded. Furthermore, we use a common form for the utility function in both specification. The two models are similar in several respects; they differ in the structure of the goods markets (competitive vs. monopolistic competitive producers) and in the rigidities imposed for monetary policy to have real effects. In the former model, monetary shocks alter the opportunity costs of hiring productive labor and therefore the production possibilities of the economy. In the latter, real effects obtain because monopolistic competitive firms accept to meet whatever level of aggregate expenditure exists in the economy at prices which are exogenously allowed to vary only in a predetermined and sluggish manner. Since the evaluations of this two types of models has been only partially carried out in the literature, our analysis may shed light on the relative performance of these two models for a common set of questions.

4.1 Limited Participation model

There is a large number of consumers maximizing a utility function, which values consumption and leisure, of the form

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

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, a budget and a capital accumulation constraint of the forms:

$$\begin{aligned} c_t p_t &\leq M_t - I_t + W_t n_t && (2) \\ M_{t+1} &= D_t + F_t + r_t p_t k_t + M_t - I_t + W_t n_t - c_t p_t - x_t p_t && (3) \\ x_t &= k_{t+1} - (1-\delta)k_t + \frac{a}{2}(\frac{k_{t+1}}{k_t} - 1)^2 k_t && (4) \end{aligned}$$

where $W_t n_t$ is nominal labor income, $r_t p_t k_t$ is nominal capital income, F_t and D_t are the dividends received from the end-of-the-period liquidation the firms and the banks and a is an adjustment cost parameter. Implicit in this formulation are three important facts: deposits I_t are chosen before shocks are realized; consumers undertake investments x_t at the end of each period facing a quadratic adjustment cost; labor income can be used to purchase consumption goods in the same period while capital income can not. 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 received. To obtain working capital they borrow funds from the banks. Their problem is

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

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 solve is

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

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 chooses interest rates according to a Taylor-type rule of the form:

$$R_t = Y_t^{\omega_1} \pi_t^{\omega_2} \epsilon_t \tag{7}$$

where Y_t is output, $\pi_t = \frac{p_t}{p_{t-1}}$ is the gross inflation rate and ϵ_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 the chosen nominal interest rate.

In this formulation (k_t, I_t, m_t) are the states of the problem where $m_t = \frac{M_t}{p_{t-1}}$; there are two types of shocks (v_t, ϵ_t) and 12 endogenous variables $(c_t, n_t, Y_t, R_t, r_t, x_t, \pi_t, i_t = \frac{I_t}{p_{t-1}}, d_t = \frac{D_t}{p_{t-1}}, b_t = \frac{B_t}{p_{t-1}}, s_t = \frac{S_t}{p_{t-1}}, w_t = \frac{W_t}{p_{t-1}})$. In addition, is it possible to compute equilibrium values of two other variables: labor productivity ($LP_t = \frac{Y_t}{n_t}$) and the slope of the nominal term structure ($SL_t = R_t^{\infty} - R_t$), where R_t^{∞} is the nominal interest rate for long maturities.

The dynamics of selected endogenous variables in response to a contractionary monetary shock for the parameter values appear 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 are risk averse, consumption declines less than output and investment must decrease to maintain the resource constraint 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. The fall in consumption produces a decline in real balances because of the cash-in-advance constraint and a liquidity effect is obtained. Finally, the slope of the term structure decreases and labor productivity temporary increases.

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 disturbance and consumption increases by a smaller amount since the shock is temporary and agents prefer to smooth consumption. The increase demand of labor 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

Since the sticky price (SP) model is to a large extent similar to the limited participation (LP) economy, we only briefly describe it and highlight differences when they emerge. There is a large number of identical consumers maximizing a utility function defined over consumption, leisure and real balances of the form

$$\max_{\{c_t,n_t,k_{t+1},M_{t+1}\}} E_0 \sum_t \beta^t \left(\frac{c_t^{\mu}(1-n_t)^{1-\mu})^{\gamma}}{\gamma} + \frac{\theta}{1-\nu} \left(\frac{M_{t+1}}{p_t}\right)^{1-\nu} \right) \tag{8}$$

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

$$M_{t+1} \leq r_t p_t k_t + W_t n_t + M_t + S_t + F_t - (x_t + c_t) p_t \tag{9}$$

and the capital accumulation constraint (4), where F_t are as before the dividends received from the end-of-the-period liquidation of the firms producing intermediate goods, $p_t = (\int_0^1 p_{it}^{1-\rho} di)^{\frac{1}{1-\rho}}$ is the CPI index and $c_t = (\int_0^1 c_{it}^{\frac{1-\rho}{\rho}} di)^{\frac{\rho}{1-\rho}}$ is the consumption bundle.

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

$$\min_{\{k_{it},n_{it}\}} (r_t k_{it} + W_t n_{it}) \tag{10}$$

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

$$\max_{\{p_{it+j}\}} \frac{U'(c_{t+1})}{p_{t+1}} \eta^j F_{t+j} \tag{11}$$

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

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

$$p_t = (\eta p_{t-1}^{1-\rho} + (1-\eta)\tilde{p}_t^{1-\rho})^{\frac{1}{1-\rho}} \tag{12}$$

where $\tilde{p}_t = (\int_0^1 \tilde{p}_{it}^{1-\rho} di)^{\frac{1}{1-\rho}}$; $\tilde{p}_{it} = \frac{\rho}{\rho-1} \frac{E_t \sum_{j=0}^{\infty} \frac{U'(c_{t+j})}{p_{t+j}} \eta^j Y_{it+j} MC_{it+j}}{\sum_{j=0}^{\infty} \frac{U'(c_{t+j})}{p_{t+j}} \eta^j Y_{it+j}}$ 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 (7) and that at that rate it stands ready to provide the money consumers demand.

In this model there are two states (k_t, m_t); two types of shocks (v_t, ϵ_t) and 10 endogenous variables $(c_t, n_t, Y_t, R_t, r_t, x_t, \pi_t, f_t = \frac{F_t}{p_t}, s_t = \frac{S_t}{p_t}, w_t = \frac{W_t}{p_t})$. 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 disturbance for the parametrization of the model reported in table 1 appear in the first column of figure 2. It is 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 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; the slope of the term structure decreases and labor productivity temporary increases. While the impact effects in the two models are qualitatively similar, the persistence of the responses 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. Note that both models generate the so-called "prize puzzle": inflation increases in response to contractionary policy shocks as a consequence of smoothing desires of the agents and the adjustment costs to capital.

The responses to an expansionary technological disturbances (second column of figure 2) are also similar to those obtained with the LP model: employment, output, capital and consumption all increase, the nominal interest rate rises 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 larger than the one obtained in the LP economy.

4.3 Evaluation

4.3.1 Robust features

To study the validity of the two prototype models we first examine whether the basic features of dynamic responses we have outlined depend on the exact features of the model. We initially focus attention on the responses of the endogenous variables to monetary disturbances since these two models differ primarily in the way these disturbances are transmitted to the real economy. Later on we discuss the robustness of the restrictions these models generate 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. The first column of figure 3 reports correlations for the basic LP model (whose responses are presented in figure 1); the second correlations when adjustment costs are smaller (a=0.5); the third correlations when the utility function is logarithmic in consumption and labor is indivisible; the fourth 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 correlations when the coefficient of relative risk aversion is large (and equal to 5) and the last column correlations when adjustment costs are extreme (a=30). The first column of figure 4 reports correlations for the basic SP model (whose responses are in figure 2), the second correlations when prices are changed more often ($\eta = 0.25$); the third correlations when the elasticity of money demand is high ($\epsilon = 20$); the fourth 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 correlations when the utility function is logarithmic in consumption and labor is indivisible and the last correlations when a money growth rule is employed in place of a Taylor rule.

Both figures indicate that the sign and, in some cases, the magnitude of several correlations is robust. In response to policy disturbances, the LP economy typically generates negative cross correlation between inflation and output, and between inflation and real balances, while the correlation between output and real balances is positive throughout the range except, perhaps at one year lead with two specifications. The cross correlations of the slope and output and of the slope and real balances are 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 correlation between inflation and output is 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 contemporaneously 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.

The last columns of each figure present situations where the shape and/or the signs of the correlations are altered. In the last column of figure 3 adjustments costs are so large that capital is constant over the adjustment path; therefore these would be the cross correlations obtained in a model without capital. In the last two columns of figure 4 the shape of cross correlations functions are altered, but the signs of the contemporaneous correlations in the first three boxes are unchanged. We have checked the signs and shapes of these cross correlation functions when parameters are varied within these ranges and with other interesting specifications, (e.g. one where steady state inflation is high and one where agents' utility value only consumption) and confirmed that they do not change in these alternative setups. Hence, sign restrictions in the cross correlation function of several pairs of variables in response to monetary shocks are robust to variations of parameters within a reasonable range, to functional forms for the primitives and, to some extent, to alternative policy rules.

4.3.2 Identification of shocks

One could use some or all of these restrictions to characterize monetary shocks in the two models. For reasons which will become apparent, we will limit attention to a subset of these constraints and select sign restrictions on the cross correlation function of output, inflation and the slope of the term structure for the LP model and on the cross correlation function of output, inflation and real balances in the SP model. One reason for choosing two different sets of variables is that technology and monetary disturbances imply different sets of sign restrictions for these three variables in the two models. Therefore, they can be jointly identified within the same VAR model. We run a five variable VAR with output, inflation, real balances, the slope of the term structure and labor productivity and call monetary those disturbances which produce the same sign restrictions on the cross correlation function of the three pairs of variables. In searching the space of identification for shocks which satisfy the restrictions, we follow Canova and De Nicoló (1999), whose approach is briefly summarized in appendix A. Figures 5 and 6 report the cross correlations obtained in response to identified monetary disturbances for three different data sets: the US, the UK and the EURO land using 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. In both cases only sign restrictions on the contemporaneous correlation function are used.

Two features of the figures deserve comments. First, while the monetary shocks we derive satisfy the sign restrictions by construction, the shapes of the cross correlation functions of the three pairs of variables in the data are somewhat different from the theoretical ones.

Second, LP sign restrictions fail to recover any monetary shock in the UK, while SP sign restrictions do not produce monetary shocks with Euroland data. That is, with these two data sets, no combination of reduced form residuals produces cross correlations for output, inflation and the slope (or real balances) with the required contemporaneous signs. Splitting the samples in various ways to account for different monetary regimes or increasing the lag length of the VAR to account for possible MA components or missing variables did not change the results. This lack of identificability, which is the first important test of the two theories, however, it is not peculiar to these two data sets and generalizes to both models and all countries 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 monetary shocks under either LP or SP restrictions with any of the three data sets. Similarly, had we imposed robust contemporaneous sign restrictions on four variables (instead of three), we would not have been able to obtain monetary shocks in any case. These results taken together suggest that both models seriously misspecify the dynamic comovements of the variables in response to shocks.

Are the responses of the variables of the system to an identified monetary shock reasonable? With both identifications schemes a contractionary monetary policy shock induces output to decline. The qualitative pattern of output responses is similar across countries and identifications: responses are instantaneous, but the through typically occurs after about 3/4 quarters indicating sluggishness in output adjustments. The qualitative pattern of inflation responses however differs across identification schemes: responses are at times negative and at times positive. Since inflation responses are positive in both models, there is at least one data set with which each model is at odds. Furthermore, there appears to be little conformity in inflation responses across countries: in US they are instantaneous and strong with both schemes, while in the other two countries they are smoother and more sluggish.

4.4 Qualitative Comparison

To evaluate the two models we next examine the dynamic responses of two other variables to identified monetary shocks: 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 we restrict attention to these two variables, considering the dynamic behavior of any other of the three variables does not create any circularity since only the sign of contemporaneous comovements has been used at the identification stage.

There are at least two reasons for why a comparison based on these variables may be informative about the economic 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 should be entertained or not in structural VARs (see e.g. Leeper and Gordon (1994)). Given that both models have built in liquidity effects, failure to produce this feature in the data will be considered an important economic failure of the models.

Second, the behavior of labor productivity has been the focus of several studies 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 explanations of economic fluctuations. The dynamics of labor productivity are very similar in the two models: since a contractionary policy shock reduces employment more than output, 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. Our task will be to examine whether labor productivity responses in the data qualitative conforms to the predictions of the models.

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), scaled so that the variance of the monetary policy innovation is the same. With the LP identification the responses of real balances in the US have the correct sign but are more pronounced than in the model. For Euro data, real balances responses have the wrong contemporaneous 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 exists in both data sets but it is very short lived relative to the one of the model There are two alternative explanations for this lack of persistence: either the movements in short term interest rates are mean reverting or strong expected inflation effects, driving long term interest rates almost immediately up, are present. Since output dynamics are inconsistent with the presence of both these features, one must conclude that SP identified monetary policy shocks have hard time to generate sufficeintly persistent liquidity effects.

Labor productivity responses obtained with the LP identification scheme differ from those of the model: both the sign and the direction of the adjustment are incorrect with both data sets. The model predicts that employment reacts more strongly than output to monetary disturbances, but the opposite is true in the data. This could be due to labor hoarding, sluggishness in employment adjustments, measurement errors, or the presence of other factors (e.g. utilization or inventories) not modelled in our framework that are more volatile than output in response to the shocks. Notice, that the smaller response of hours to monetary shocks is not due to the institutional differences of labor markets or to different rigidities present in the two continents: all data sets give the same result. With the SP identification the data is more supportive of the model: for the US the sign of instantaneous labor productivity responses in the data is correct even though its dynamic shape is not. For the Euro data the responses are remarkably similar to the one obtained in the model.

To summarize, both models missrepresent the dynamics present in 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, one data set and this failure generalizes when a larger number of sign restrictions are considered. Second, the LP identification scheme 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. Third, with the SP identification scheme monetary shocks generate instantaneous responses of the slope of the term structure which are very small and lack the persistence present in the data.

4.5 Quantitative Comparison

Although the two theories produce dynamics which are qualitatively at odds with the data - and this would be sufficient to cast doubts on their use for policy and economic analyses - we proceed to undertake a quantitatively evaluation of the two models to highlight some additional interesting properties. There are many ways to quantitatively assess the discrepancy between the model and the data and a number of variables for which the exercise could be performed. Here we focus on two variables and three statistics which provide useful economic information to potential users of these models. Our aim is not to be exhaustive: we only attempt to assess the properties of the models along a number of interesting dimensions.

Both academics and policymakes have been interested for decades in measuring the real effects of monetary policy. Chari, Kehoe and McGrattan (2000) have argued that simple SP models are incapable to generate persistence in output responses suggesting that price rigidities alone are insufficient to propagate monetary disturbances. Figure 2 confirms that in this type of models the real effects of monetary shocks die out very quickly. One question of interest is whether the output effects obtained in the data restricted with SP sign restrictions quantitatively resemble those of the model and whether the LP model offers a more appropriate description of the data. To provide evidence on this issue, we present in table 2 the value of the half life of output responses in the model and the 68% Monte Carlo band for the estimates obtained in the data when the two types of sign restrictions are imposed. Intuitively, whenever the bands do not include the theoretical value, the data and the model are at odds which each other. Furthermore, we measure the percentage of the variance of output at 4, 8 and 24 steps horizons due monetary innovations. This statistics is typically used by VAR researchers to assess the contribution of monetary disturbances to business cycle fluctuations. Estimates have varied quite a lot in the literature. We go from 0-20% in studies like Uhlig (1999); to 15-30% in studies like Gordon and Leeper (1994) to values of 40% or above in studies like Canova and De Nicoló (1999). To make the comparison appropriate, since there are five shocks in the VAR and only two shocks in the theory, we augment the 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. In table 3 we present 68% Monte Carlo bands for these percentages in each of the data sets. Once again, when the model values are outside the 68% range, the conformity between the data and the model is low.

A succesful monetary model must also account for the link between money and inflation.

Lucas (1996), for example, argues that in the long run the correlation between the two is almost perfect. To quantitatively measure this link in our case we compute the cumulative effect of monetary shocks on inflation and calculate the probability that the value obtained in the two models is lower than the one found in the data. Table 4 reports these numbers at horizons 4, 16, ∞: a value smaller than 16% or larger than 84% (2.5-97.5%) indicates that the cumulative response of inflation produced by the model is different from the one found in the data by one (two) standard deviations.

As expected the half life of output responses in SP restricted data is much longer than in the model: for example, the median half life is 10 quarters in US data and 17 quarters in UK data and for this data set, the lower bound of the 68% band is 6 quarters (the value in the model is 1). Therefore, we confirm that the persistence of output responses in the SP model is different from the one in the data. For the LP scheme, the half life of 9 quarters produced by the model is close to the median for the US data (the value is 11), while is too short for Euro data where the output responses are humped shaped and very persistent.

The percentage of output variance accounted for by monetary shocks in US at the 24 step horizon is between 11 and 43% with the LP scheme and 3 and 34% with the SP scheme. For Euroland output the range is large (from 0.1 to 53%) and in the UK SP-based monetary shocks account for 9-31% of output variance. In comparison, and regardless of the exact parametrization of the variance of different shocks, monetary disturbances in the two models account for less than 1% of output variance. Hence, both models lack internal propagation mechanism and fail to account for the size of output responses to monetary shocks.

Cumulative inflation responses for the LP model in the long run are very near the median value of the estimated band and at no horizon the cumulative response of the model falls outside the 95% range for both data set. For the SP model cumulative inflation responses consistently have the wrong sign for both data sets and the probabilistic measures confirm this. Note also that the uncertainty present in estimating cumulative inflation responses in the data is not sufficient to reconcile differences between the SP model and the data. Hence, long run inflation dynamics in the SP model are inconsistent with those of the data.

4.6 Other implications of the models

The analysis so far has focused attention on monetary disturbances because the two models differ primarily in the way monetary disturbances affect the real side of the economy. However, our approach has not been tailored to the identification of monetary disturbances and can be used to compare the models and the data for any disturbances for which robust restrictions exist. To illustrate this point we examine the relative performance of the two models when the sign restrictions on the responses to technology disturbances are used to identify these shocks in the data.

As with monetary shocks, there are several robust restrictions one could use. For the sake of symmetry, we focus once again on 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 attempt to simultaneously identify the monetary shocks we have discussed in the previous subsections and technology disturbances. Figure 9 reports the pairwise cross correlation functions of these variables in response to technological disturbances in the models, figures 10 and 11 the pairwise cross correlation functions obtained in the data once technology shocks are identified using contemporaneous sign restrictions. Figures 12 and 13 report impulse responses of the five variables to the identified shocks. Also in this case, imposing the whole vector of sign restrictions results in outright rejections of both models.

A few features of the figures deserve discussion. First, no technology shock is identifiable 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 the two models have very similar implications for these three variables in response to technological disturbances (see figures 1 and 2) it is comforting to find these similarities. Clear differences between the predictions of the models and the data however emerge when we consider the other two variables: real balances always decline after a positive technology in the data with the SP scheme while with the LP scheme they decrease in the US. A negative response of real balances is counterfactual and highlights a substantial mismatch between the theory and the data [1]. Labor productivity responses have signs which depend on the identification scheme and the country. Since in the models labor productivity responses were always positive, both schemes fail to account for the behavior of this variable with at least one data set.

In both models technology shocks are the dominant source of real fluctuations. One may be curious as to whether the data conforms to this prediction. The answer is positive: the value obtained in the models at the 24 step horizon is within one standard error band of the percentage found in all data sets with all identification schemes.

4.7 Improving the match between the model and the data

Both models display important shortcomings in explaining the dynamic response of a number of variables. One may be therefore interested in going back to the drawing board and examine whether there are changes in their structure which may improve the quality of the approximation. In this section we concentrate attention on two particular extensions of the setup we have used so far and study whether the inclusion of these features significantly improves the qualitative match of the LP model to the data.

We have seen that the dynamics of labor productivity in response to monetary disturbances when the data is restricted with the LP scheme are inconsistent with those of the model. We have also argued that this occurs because hours move more than output in the data and this is not the case in the model. Here we examine whether the introduction of

[1] It is however possible that the odd behaviour 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.

factor utilization helps to reconcile this discrepancy.

Capacity utilization is a natural candidate to consider since is has been found important in making the dynamics of DSGE models more coherent with the data (see Burnside and Eichenbaum (1996) or Neiss and Pappa (2001)). We add capacity utilization to the 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 principle, if variations in capacity utilization are limited, it may be possible to increase the impact of monetary shocks on employment and therefore make labor productivity fall in response to monetary shocks. We assume that varying capacity utilization 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 produce the sign restrictions we have used to identify monetary shocks for a range of values of the parameter which controls the depreciation cost due to utilization. Second, we searched within this range for the value of the parameter which gave us a negative response of labor productivity. We found none. In figure 7 we show what happens to labor productivity when this parameter is fixed to the standard value of 1.56: responses are more sluggish relative to the baseline case, as intuition would suggest, but capacity utilization alone is not sufficient to change the relative magnitude of output and hours responses.

Next, we split employment decisions into two margins: hours worked (intensive) and number of people working (extensive) (as in Hansen and Sargent (1988)). We assume that there is sluggishness in employment 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 hours in response to the shocks but not employment, at least instantaneously. We treat labor as we have treated capital previously and summarize the alterations of the model produced by this feature in appendix B. The sign of the dynamics of output, inflation and the slope are qualitatively unchanged when we allow this new feature in the model. The responses of labor productivity to monetary shocks are also reported in figure 7: allowing effort and employment to be separately chosen does not change the sign of labor productivity but produces humped shaped responses as we can observe in the data.

In conclusion, none of the two modifications makes labor productivity fall in response to policy disturbances. Hence, a major adjustment of the model is necessary if labor productivity dynamics matter for the useres of this theory.

4.8 Policy Analyses

Based on the evidence we have accumulated, it is safe to say that both models are unsuited for conducting credible policy analyses. 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 policy shock under three different rules: a Taylor rule (TR), a partial accommodative rule (PA), linking interest rates to real balances, and an inflation targeting rule (IR). Although no claims of optimality is made, our exercise can highlight the relative welfare ranking of these three policies. For comparisons we calculate the compensating variation of consumption needed for agents to remain on average on their steady state path. In other words we compute:

$$\bar{U}_1 = \sum_t \beta^t \frac{(c_t^j - \bar{c}_{LP}^j)^\mu (1 - n_t^j)^{1-\mu})^\gamma}{\gamma} \qquad j = TR, PA, IR \tag{13}$$

$$\bar{U}_2 = \sum_t \beta^t \frac{(c_t^j - \bar{c}_{SP}^j)^\mu (1 - n_t^j)^{1-\mu})^\gamma}{\gamma} + \frac{\theta}{1-\nu} (\frac{M_{t+1}}{p_t})^{1-\nu}) \tag{14}$$

where $j = TR, PA, IR$, and $\bar{U}_i$ are the steady state utilities in model $i = LP, SP$. A policy rule is preferable if it requires smaller (in absolute value) compensating variations in consumption. When we choose standard parameter values ($\beta = 1.04^{-0.25}$, $\gamma = -2.0$, $\mu = 0.5$, $\theta = 1.0$, $\nu = 7$) the ranking of policies is the same for both models: partial accommodative rules are best (0.17% and 0.03% of steady state consumption); Taylor rules follow very closely (0.18% and 0.04% of steady state consumption); and inflation targeting rules are distant last (0.66% and 0.15% of steady state consumption). The relative ordering remains when we vary $\gamma \in [-5.0, -1.0]$; $\mu \in [0.1, 1.0]; \nu \in [2, 20]$. Not surprisingly, the numbers obtained with the first two rules are not statistically different. The intuition behind the slight preferability of partial accommodative rules is simple: consumption and employment are less volatile with this rule because in response to shocks real balances responses are much smoother than either output or inflation responses. Notice that, the more aggressive is the Taylor rule, the worse is the welfare outcome.

5 Conclusions

This paper has provided an alternative methodology to compare DSGE models to the data. The suggested procedure is simple, easily reproducible, and avoids some of the problems encountered with existing approaches. It also allows to evaluate models in response to different shocks and to respecify the theoretical construction at any level of the evaluation when the discrepancy with the data is large. Furthermore, it is designed to provide both qualitative and quantitative comparisons and is suited to answers a number of economic questions 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. It also takes the view that economic (as opposed to statistical) validation is what matters for users of DSGE models and we propose an approach which takes these concerns into account. Finally, the procedure takes care of 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 which hold regardless of the exact parametrization, of particular functional choices or of certain features that a researcher may want to include in the specification. Once robust implications are discovered, we used them as identification devices to construct structural shocks in VAR models. That is, we calibrate the model to the data using a minimal set of robust restrictions derived from a theory. We then employ responses to identified shocks in the data and in the model to provide an economic metric to compare the two 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.

We use this machinery to evaluate two, by now standard, DSGE models of money against the data. We show that the models are at odds with the data in several ways. First, for some countries the robust restrictions that these models deliver have no counterpart in the data. Second, the qualitatively dynamics of certain variables 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 generates a dynamic shape for the slope of the term structure which are not fully consistent with those present in the data. Third, quantitatively speaking, the match between the models and the data is poor. For example, the magnitude of output responses to monetary disturbances in both models are small relative those found in the data. In addition, in response to policy shocks, the sticky price model produces real effects which lack the persistence found in the data (see Chari, Kehoe and McGrattan (2000)) and it is unable to match inflation dynamics that result from these shocks. Fourth, dynamics in response to certain shocks are fragile, in teh sense that they depend on the data set used.

We show that two simple alterations of the basic the limited participation structure are unable to bring the responses of labor productivity in line with those observed in the data, suggesting that more drastic modifications are needed to make the theory consistent with the data in this dimension. Finally we demonstrate that, although the mechanics of the propagation of shocks are different, the two models have similar implications regarding the choice of particular policy rules: consumer's welfare is highest when nominal interest rate systematically responds to real balances.

Overall, our exercise has shed new light on the nature of these models and on their match with the data. This information is useful for theorists engaged in respecifying the less appealing features of their models and to policymakers, who have to decide which transmission mechanism they should believe in when important policy decisions have to be taken.

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 $P = \prod_{m,n} Q_{m,n}(\theta)$ where $Q_{m,n}(\theta)$ are rotation matrices of the form:

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

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

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

In our case Σ is the covariance matrix of reduced form VAR disturbances and the second result holds because $Q_{m,n}$ are orthonormal. The algorithm we design works as follows:
• construct matrices P and D, check if for 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 $P_{m,n}(\theta) = P * Q_{m,n}(\theta)$ for some m, n and some θ 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 θ or another θ for any pair m, n. In a system with s variables there are (s(s-1)/2) pairwise rotations, (s(s-1)/4) combinations of pairwise rotations, etc. for any given θ. So for example in a system with s=5 variables, there are 10 pairwise rotations. Note that, each pairwise rotation "decouple" the matrix of eigenvector in a particular direction. To select θ we grid the interval $[0, \frac{\pi}{2}]$ into M points. In practice, we construct $M * (s(s-1)/2)$ orthogonal decompositions of Σ, and store all those which satisfy the required sign restrictions. It is typically the case that there is a range of θ which satisfy the sign restrictions for a given (m, n). Since the impulse responses within this range are similar, we treat this range as a point. If there are different ranges which satisfy the restrictions or different combinations m, n for one θ we choose the range or the pair which maximizes the number of interpretable shocks. Canova and De Nicoló (1999) provide further steps to narrow down the class of admissible decompositions in case this was not sufficient to select a unique candidate. When there is more than one shock which

satisfies the restriction we choose the one which produces responses which have 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. More details are in Canova and De Nicoló (1999).

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 ϕ controls the costs of overutilization in terms of depreciation (it is the elasticity of depreciation to utilization). Note that for ϕ 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 hours and H_t employment. We modify the utility function of the representative agent to be

$$\max_{\{c_t,e_t,n_t,k_{t+1},I_t\}} E_0 \sum_t \beta^t \frac{(c_t^\mu(1-e_t)^{1-\mu-\theta}H_{t-1}^\theta)^\gamma}{\gamma} \tag{15}$$

where θ 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 hours instantaneously respond to shocks. The constraints faced by the consumers are now:

$$c_t p_t \leq M_t - I_t + W_t H_{t-1} e_t \tag{16}$$

$$M_{t+1} = D_t + F_t + r_t p_t k_t + M_t - I_t + W_t H_{t-1} e_t - c_t p_t - x_t p_t \tag{17}$$

$$x_t = k_{t+1} - (1-\delta)k_t + \frac{a}{2}(\frac{k_{t+1}}{k_t} - 1)^2 k_t \tag{18}$$

References

[1] Altug, S. , 1989, " Time to Build and Aggregate Fluctuations: Some New Evidence, *International Economic Review*, 30, 889-920.

[2] Broze, L , Dridi, R.. and Renault, E., 1999, Calibration of structural models by semi-parametric Indirect inference, CREST working paper 96.

[3] Burnside, C and Eichenbaum M., 1996, " Factor Hoarding and the Propagation of Shocks", *American Economic Review* 86, 1154-1174.

[4] Canova, F., 1994, "Statistical Inference in Calibrated Models", *Journal of Applied Econometrics*, 9, s123-s144.

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

[6] Canova, F. and Marrinan, J, 1996, "Reconciling the Term Structure of Interest Rates with a consumption based ICAP Model", *Journal of Economic Dynamics and Control* , 20, 709-739.

[7] Canova, F. and De Nicoló, G., 1999 " Monetary disturbances matter for business fluctuations in the G-7", forthcoming, *Journal of Monetary Economics*.

[8] Canova, F., and Pina, J., 2000, " Monetary Policy Misspecification in VAR models", CEPR Working Paper 2333.

[9] Christiano, L. and Eichenbaum, M., 1992, "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations", *American Economic Review*, 82, 430-450.

[10] Christiano, L., Eichenbaum, M., and Evans, C. 1997, "Sticky Prices and Limited Participation Models of Money: A Comparison", *European Economic Review*, 41, 1201-1249.

[11] Chari, V.V., Kehoe, P. and Mc Grattan, E., 2000, " Sticky Price Models of the Business Cycle:Can the Contract Multiplier solve the persistence problem?", *Econometrica*, 68, 1151-1197.

[12] DeJong, D., Ingram, B. and Whitemann, C., 1996, "Beyond Calibration", *Journal of Business and Economic Statistics*, 14, 1-10.

[13] Diebold, F., Ohanian, L. and Berkowitz, J., 1998, "Dynamics General Equilibrium Economies: A Framework for Comparing Models and Data", *Review of Economic Studies*, 68, 433-451.

[14] Hansen, G. and Sargent, T. (1998) " Straight time and Over time in General Equilibrium" *Journal of Monetary Economics*, 21, 213-233.

[15] Gali, J., 1999, " Technology, Employment and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89, 249-271.

[16] Gallant, R. and Tauchen, G., 1996, "Which Moment to match", *Econometric Theory*, 657-681.

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

[18] Gordon, S. and E. Leeper, 1994, “The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification”, *Journal of Political Economy*, 102, 1228-47.

[19] Gregory, A. and Smith, G., 1991, “Calibration as Estimation”, *Econometrics Reviews*, 9, 57-89.

[20] Gregory, A. and Smith, G., 1991, ”Calibration as Testing: Inference in Simulated Macroeconomic Models”, *Journal of Business and Economic Statistics*, 9, 297-303.

[21] Ingram, B., Kocherlakota, and Savin, G. 1994, ”Explaining Business Cycles”, *Journal of Monetary Economics*,34, 415-428.

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

[23] Ireland, P. , 2001, Money’s Role in Monetary Business Cycle”, Boston College, manuscript.

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

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

[26] King, R. and Wolman, A., 1999, ” What should monetary policy do when price s are sticky ” , in J.B. Taylor (ed.) *Monetary Policy Rules*, University of Chicago Press, 349-398.

[27] Leeper, E. and Sims, C, 1994, ”Toward A Modern Macroeconomic Model Usable for Policy Analyses, in Rotemberg, J. and Fisher, S. (ed), *NBER Macroeconomic Annual*, 81-118.

[28] Leeper, E. and Zha, T, 2000, ”Assessing Simple Policy Rules: A view from a complete macro model ”, Federal Reserve of Atlanta, working paper .

[29] Lucas, R., 1996, ” Monetary Neutrality”, *Journal of Political Economy*, 104, 661-682.

[30] Neiss, K. and Pappa, E, 2001, ”A Monetary Model of Factor Utilization”, Bank of England, working paper 154.

[31] Pagan, A., 1994 ”Calibration and Econometric Research”, *Journal of Applied econometrics*, 9, S1-S10.

[32] Smets, F. and Wouters, R., 2001 ” Monetary Policy in a Estimated DSGE model for the Euro Area”, European Central Bank, manuscript.

[33] Sargent, T., 1978, ”Estimation of Dynamic Labor Demand Schedules under Rational Expectations, *Journal of Political Economy*, 86, 1009-1044.

[34] Schorfheide, F., 2000, A Unified Econometric Framework for the evaluation of DSGE Models, *Journal of Applied Econometrics*.

[35] Watson, M., 1993, ”Measures of Fit for Calibrated Models”, *Journal of Political Economy*, 101, 1011-1041.

Table 1
Parameter Values

Common Parameters							
β	α	δ	γ	μ	ν_1	ν_2	a
$1.04^{-\frac{1}{4}}$	0.36	0.025	-2.0	0.5	0.2	1.6	2
N_s	π_s	C/Y	K/Y	ρ_v	ρ_ξ		
0.33	1.005	0.7	2.5	0.9	0.9		
Limited Participation Parameters							
m/c	i/c	i/b	wn/c				
1.1	0.1	0.8	0.5				
Sticky Prices Parameters							
ν	η	θ					
7	0.75	1.0					

Standard deviation of various shocks					
	technology	policy	government	preference	money demand
US	0.0072	0.0083	0.0146	0.0090	0.0054
UK	0.007	0.01	0.0112	0.0090	0.0066
Euro	0.103	0.023	0.01	0.009	0.061

Notes: US values are obtained from Leeper and Sims (1994) and Ireland (2001); those Euro land values from Smets and Wouters (2001); those for the UK inferred by the author using a variety of Bank of England publications.

Table 2
Half life of output responses

Model	Value	**US**	**UK**	**Euro**
LP	9	[1,12,40]		[40,40,40]
SP	1	[2,10,40]	[6,17,40]	

Notes: In each box 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

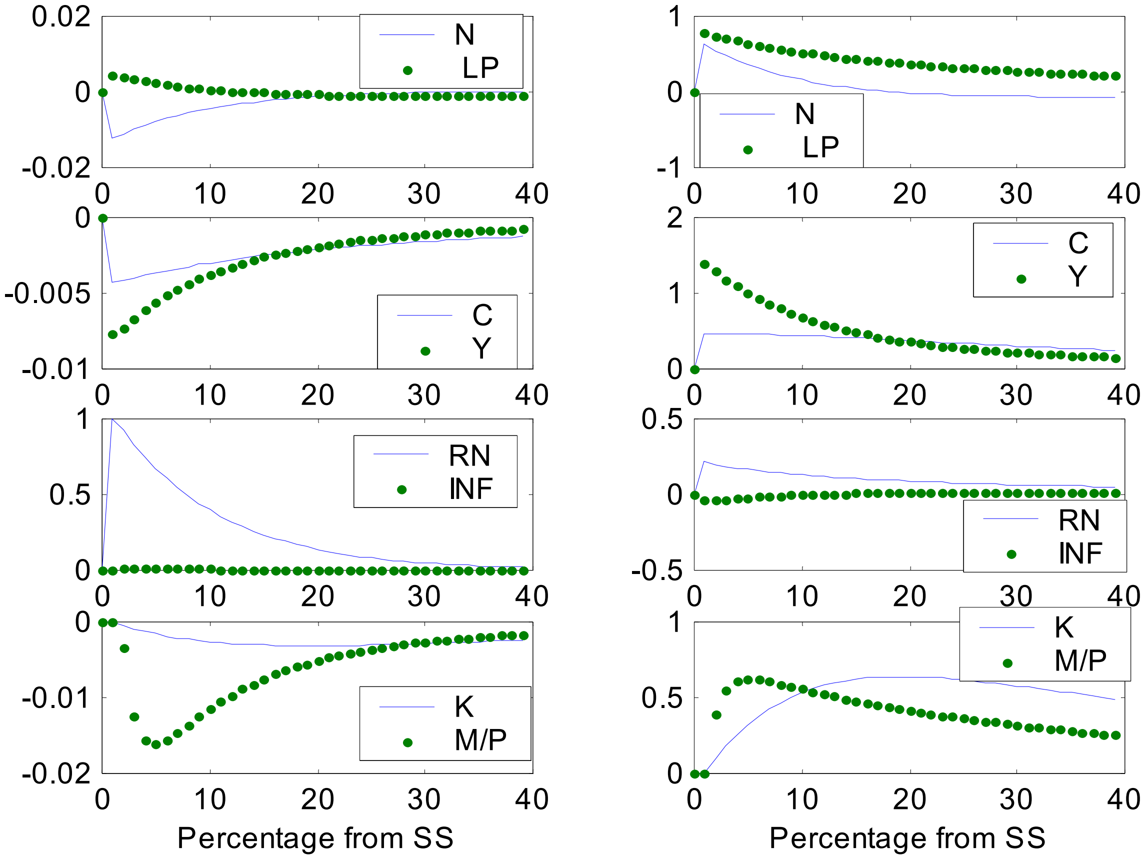
Step	**Model**	**US**	**UK**	**Euro**
		Limited Participation economy		
4	0.01	[0.41, 0.45]		[0.001, 0.13]
8	0.01	[0.30, 0.045]		[0.001, 0.34]
24	0.01	[0.11, 0.44]		[0.006, 0.52]
		Sticky Price economy		
4	0.01	[0.05, 0.34]	[0.20, 0.36]	
8	0.01	[0.02, 0.34]	[0.14, 0.35]	
24	0.01	[0.03, 0.33]	[0.09, 0.31]	

Notes: In each box 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

Step	**US**	**UK**	**Euro**
		Limited Participation economy	
4	0.92		0.46
16	0.70		0.49
∞	0.54		0.49
		Sticky Price economy	
4	1.00	1.00	
16	0.95	0.99	
∞	0.70	0.90	

Notes: Each cell reports the probability that cumulative responses of inflation in the model are lower than in the data.



0.02
0
-0.02
N
LP
0
10
20
30
40
0
-0.005
-0.01
C
Y
1
0.5
0
RN
INF
0
-0.01
-0.02
K
M/P
Percentage from SS
1
0
-1
2
0.5
-0.5
Percentage from SS

Figure 1:

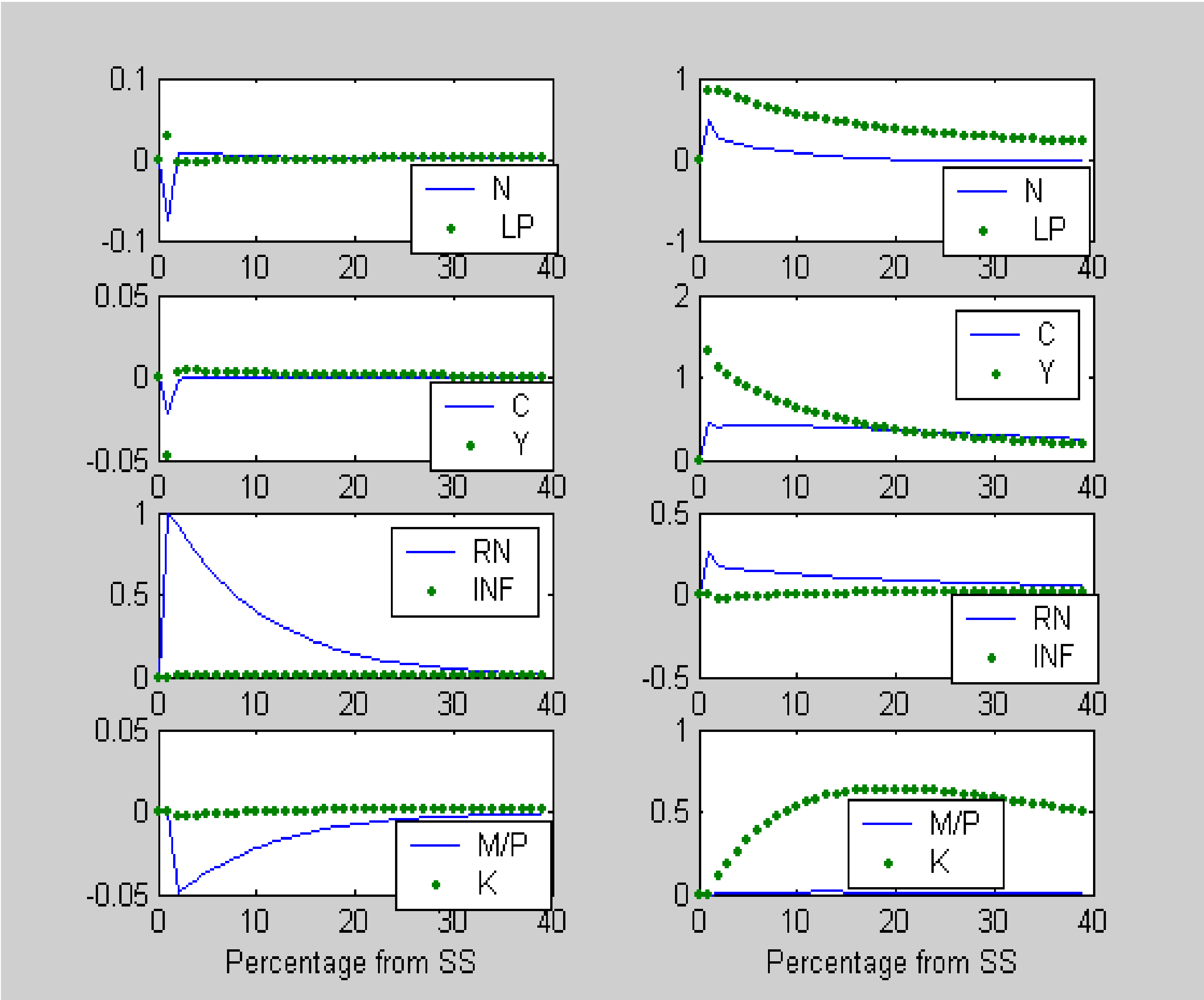


Figure 2:

Cross Correlations,monetary shock; Model LP

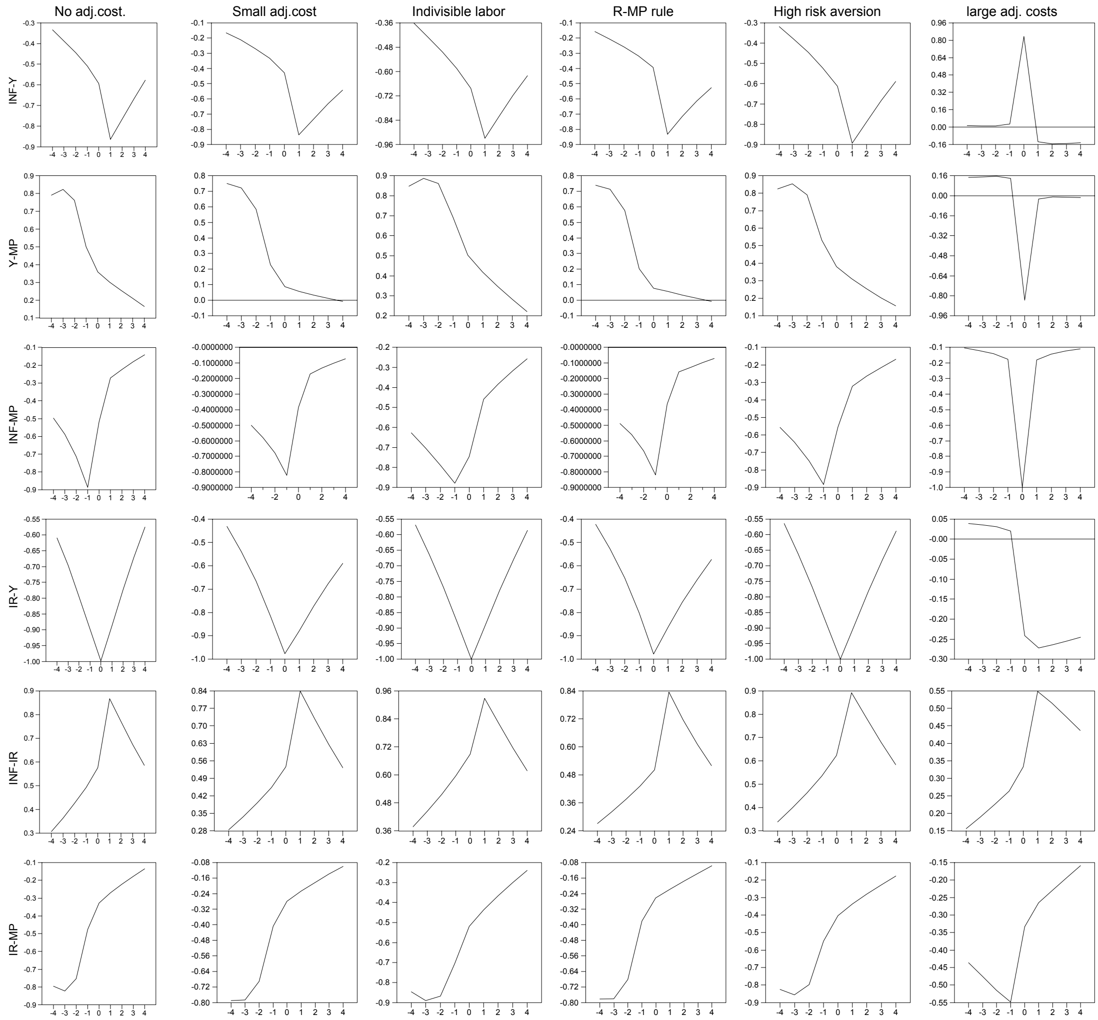


Figure 3:

Cross Correlations,monetary shock; Model SP

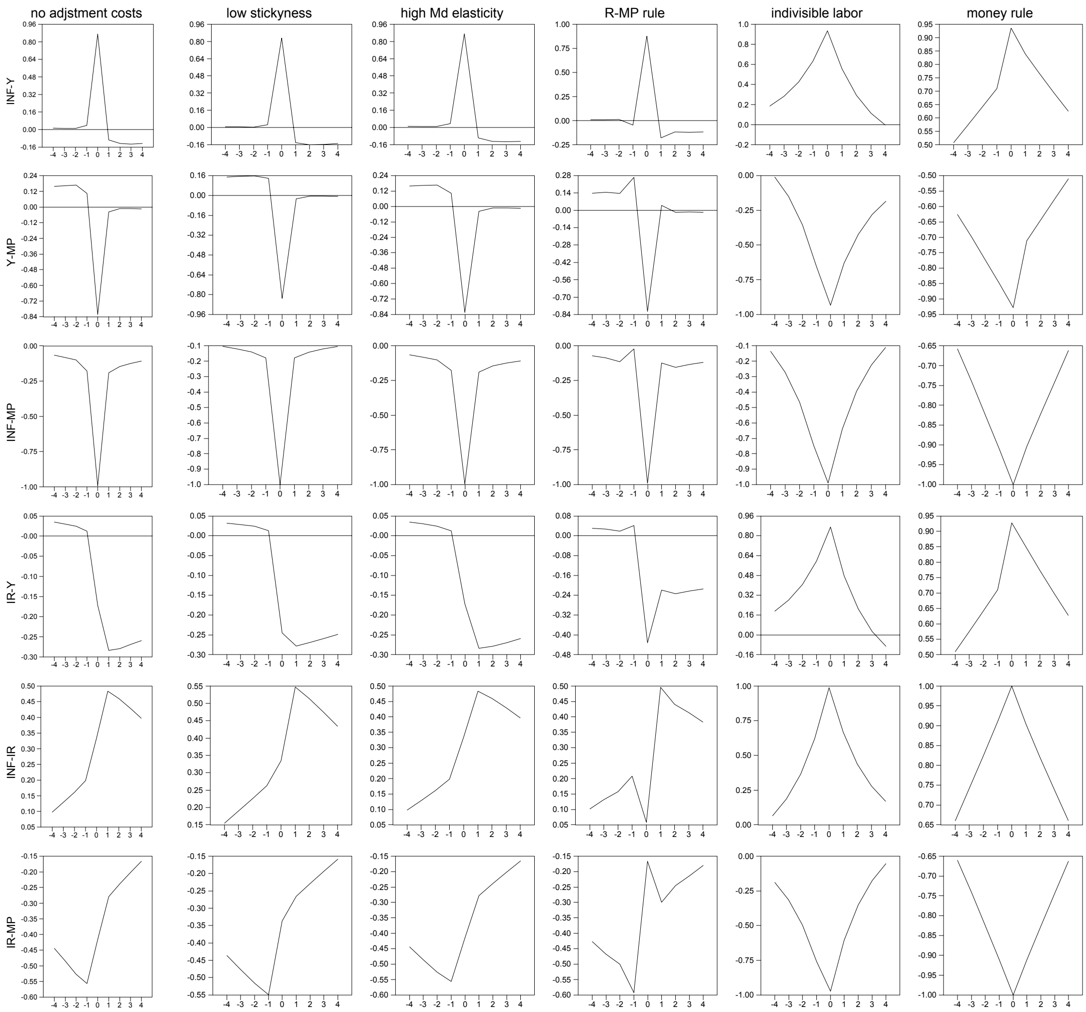


Figure 4:

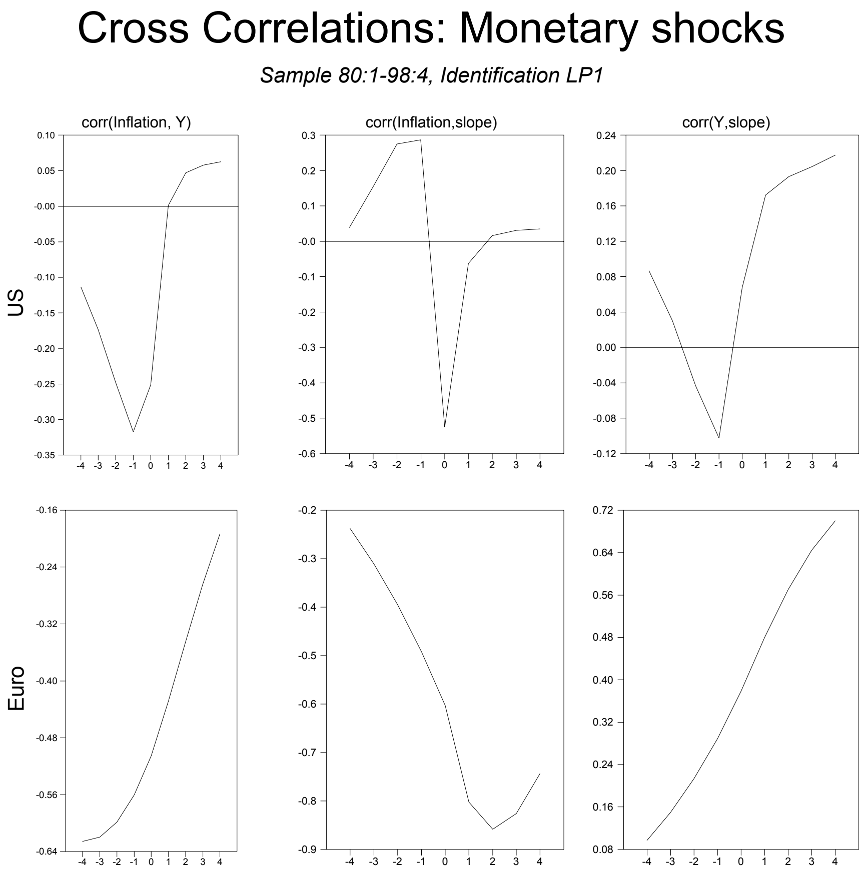

Cross Correlations: Monetary shocks
Sample 80:1-98:4, Identification LP1
corr(Inflation, Y)
corr(Inflation,slope)
corr(Y,slope)
US
Euro

Figure 5:

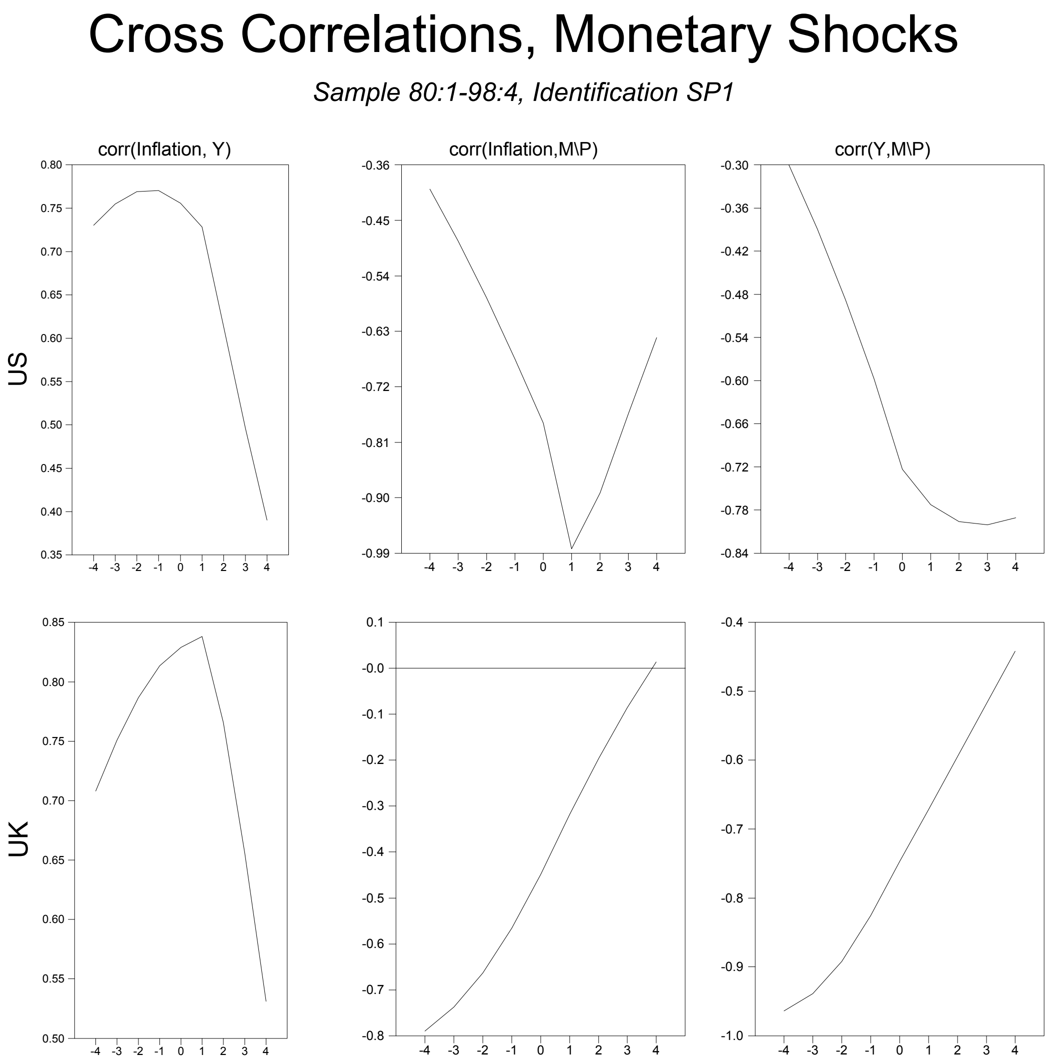

Cross Correlations, Monetary Shocks
Sample 80:1-98:4, Identification SP1
corr(Inflation, Y)
corr(Inflation,M\P)
corr(Y,M\P)
US
UK

Figure 6:

Responses to Monetary Shocks

Sample 80:1-98:4, Identification LP1

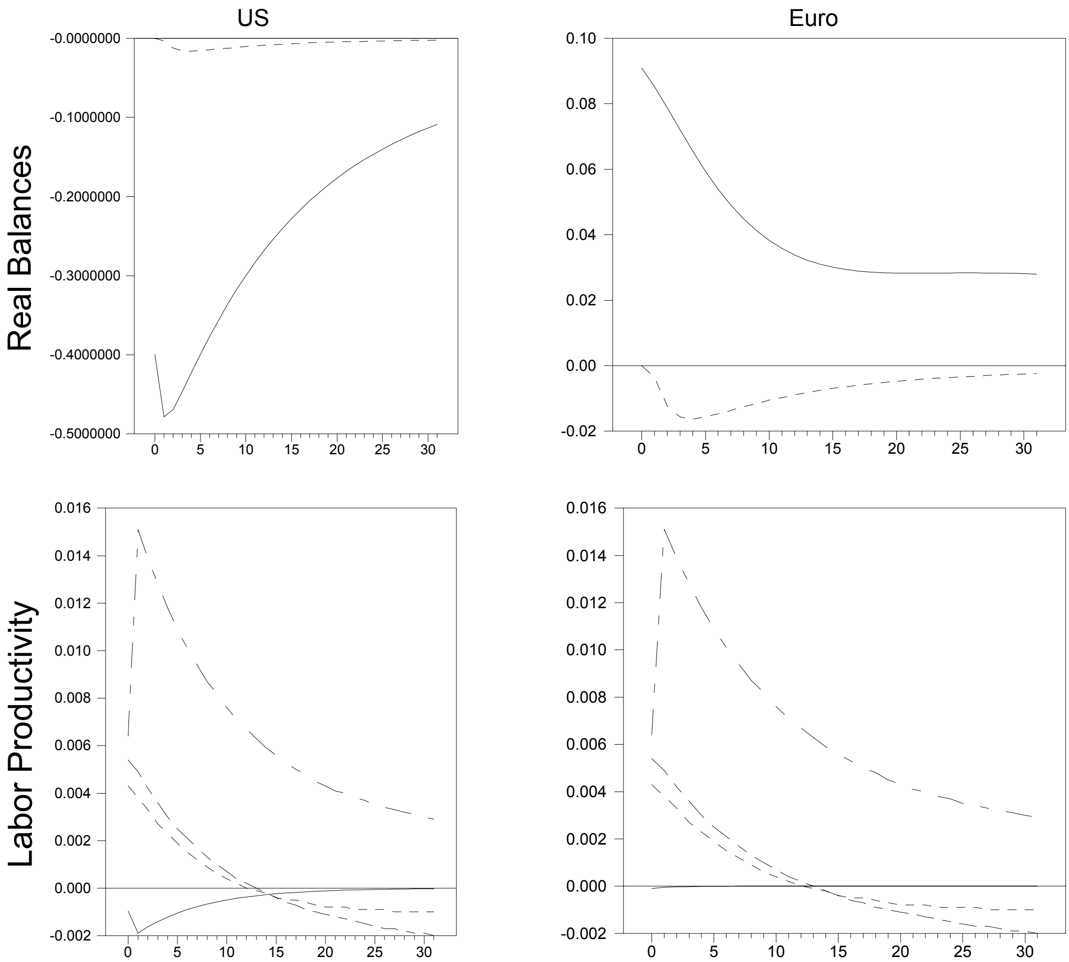


Figure 7:

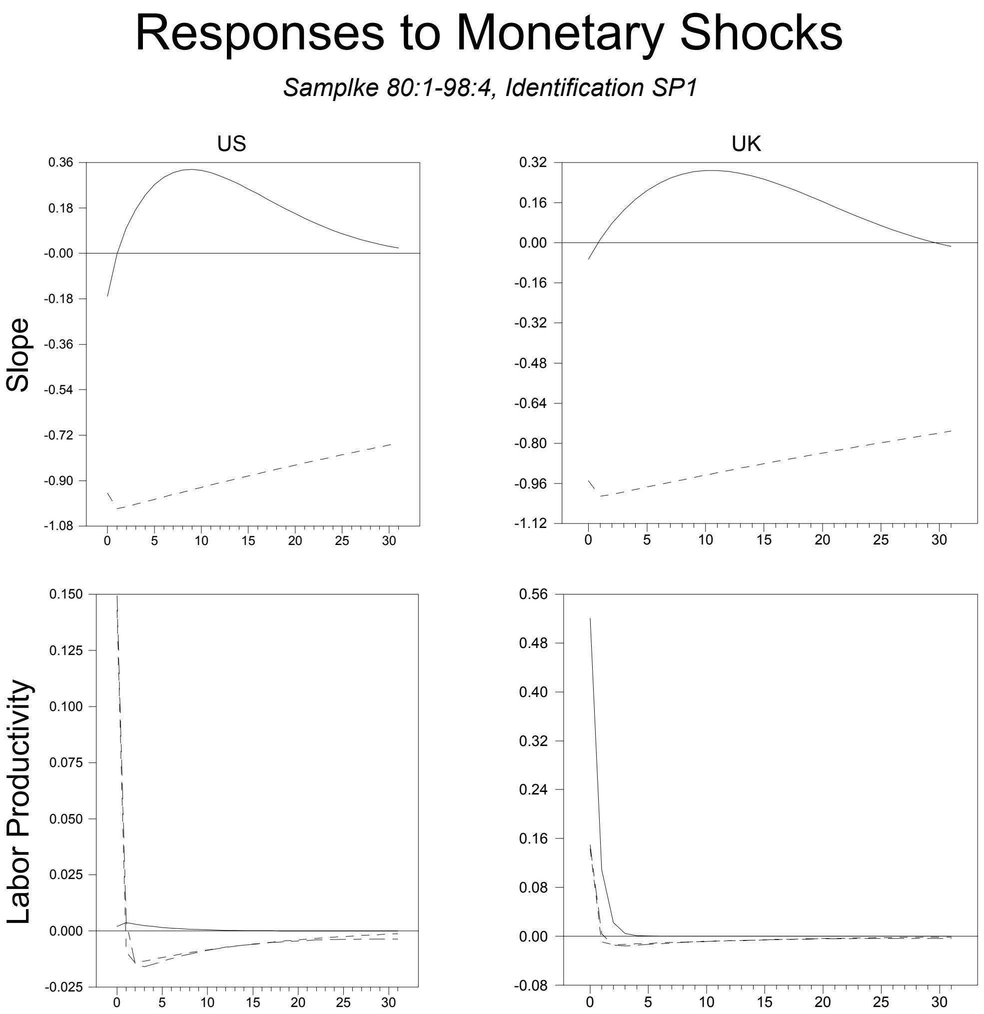

Responses to Monetary Shocks
Samplke 80:1-98:4, Identification SP1
US
UK
Slope
Labor Productivity
0.36
0.18
-0.00
-0.18
-0.36
-0.54
-0.72
-0.90
-1.08
0.32
0.16
0.00
-0.16
-0.32
-0.48
-0.64
-0.80
-0.96
-1.12
0.150
0.125
0.100
0.075
0.050
0.025
0.000
-0.025
0.56
0.48
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
0
5
10
15
20
25
30

Figure 8:

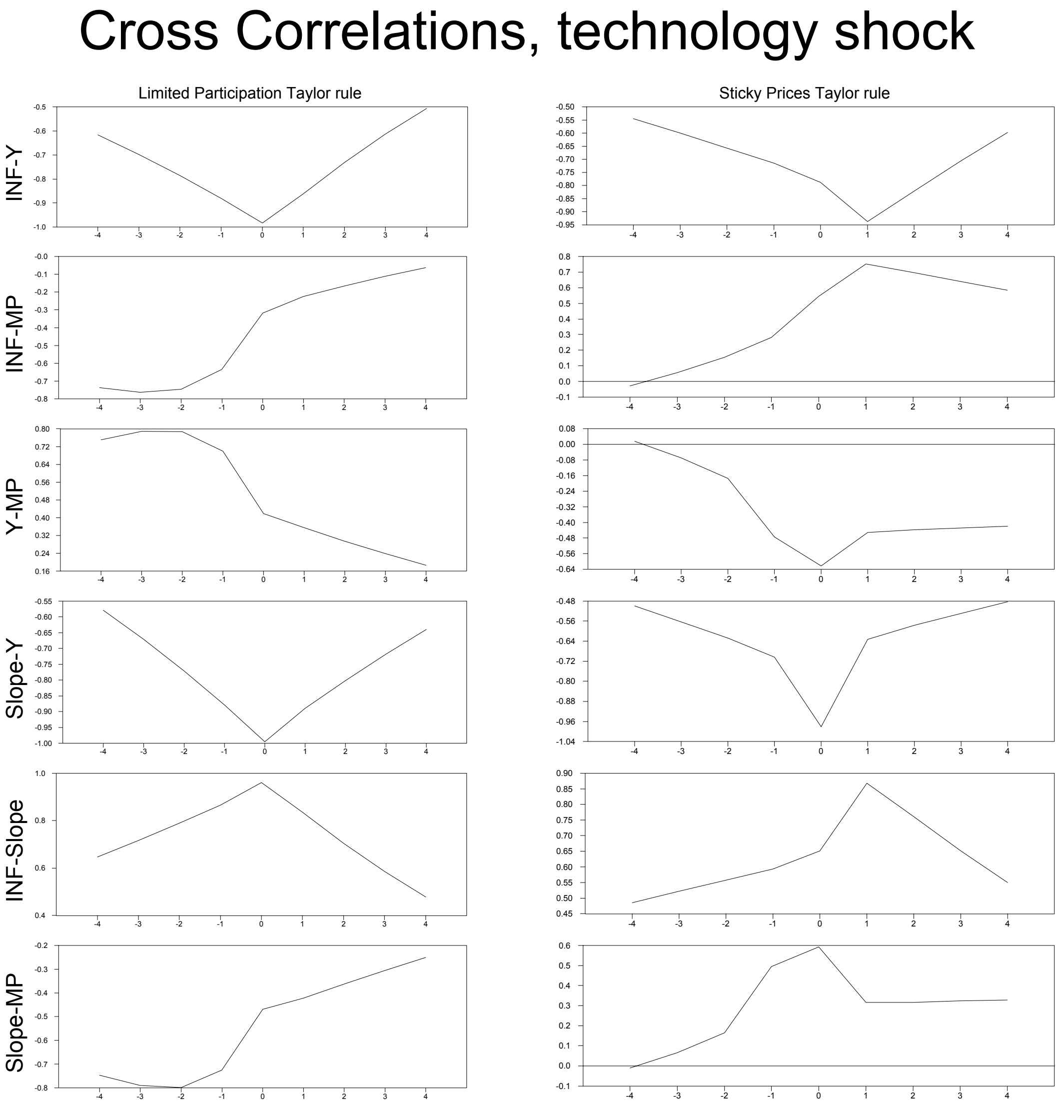

Cross Correlations, technology shock
Limited Participation Taylor rule
Sticky Prices Taylor rule
INF-Y
INF-MP
Y-MP
Slope-Y
INF-Slope
Slope-MP

Figure 9:

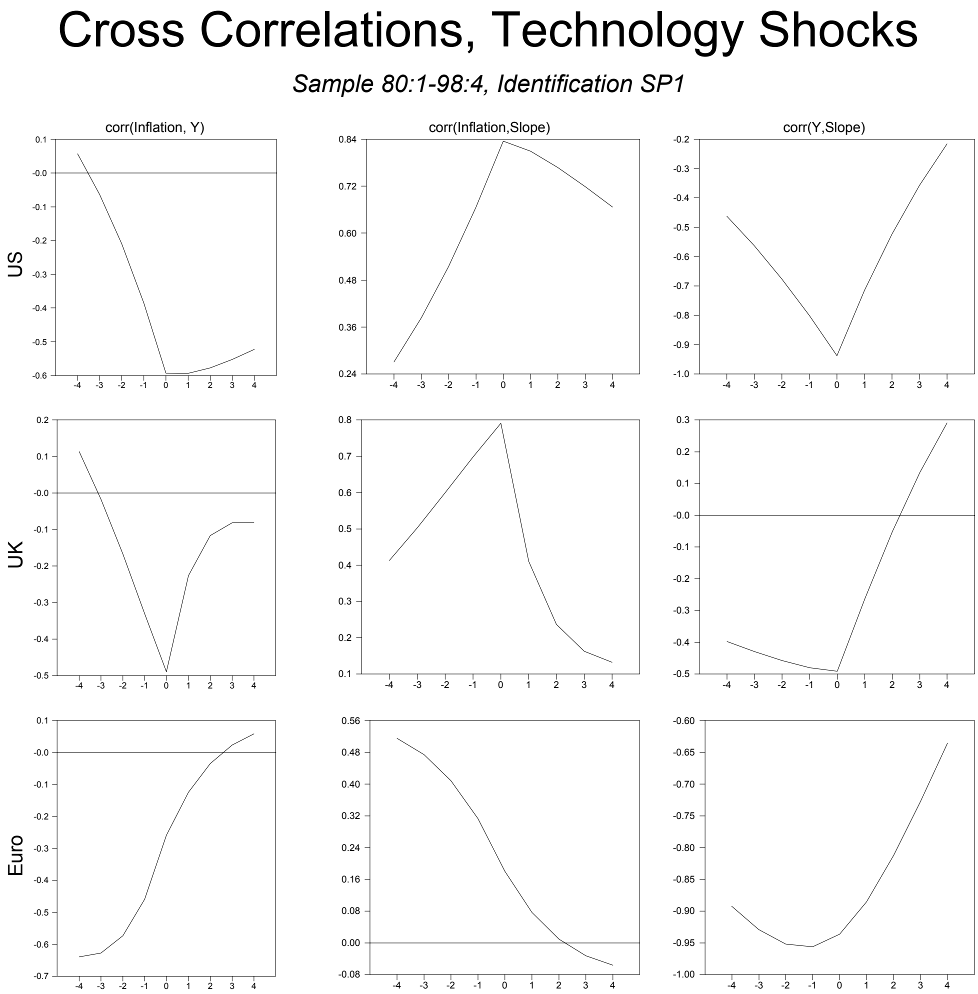

Cross Correlations, Technology Shocks
Sample 80:1-98:4, Identification SP1
corr(Inflation, Y)
corr(Inflation,Slope)
corr(Y,Slope)
US
UK
Euro

Figure 10:

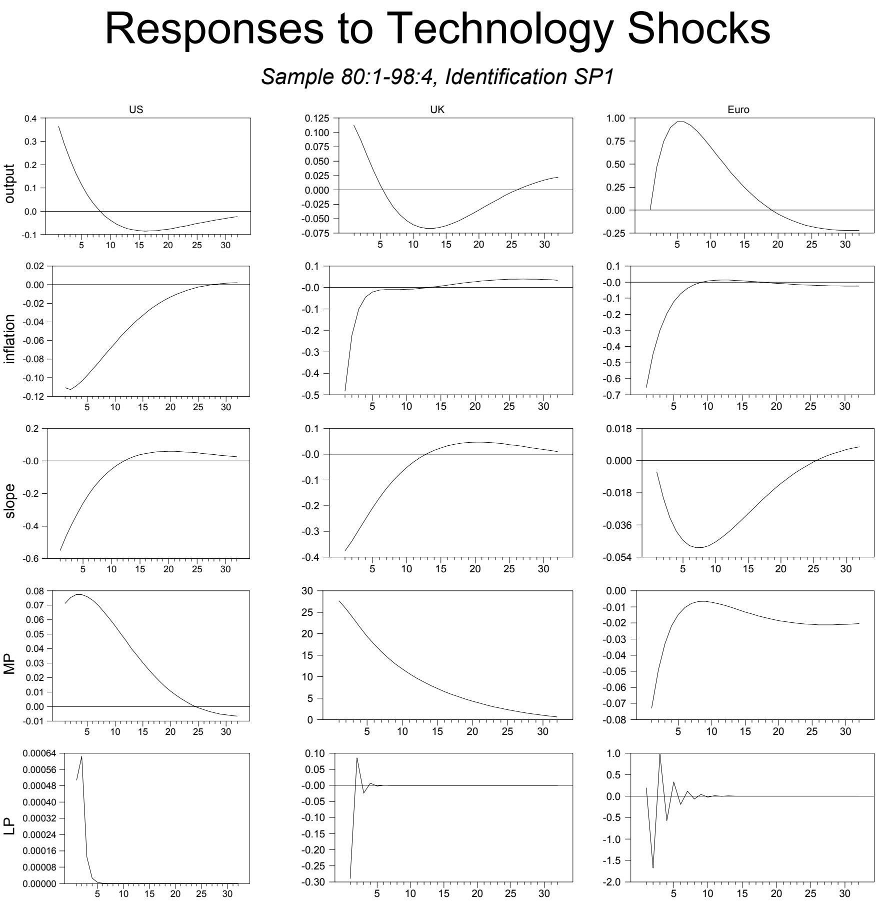

Responses to Technology Shocks
Sample 80:1-98:4, Identification SP1
US
UK
Euro
output
inflation
slope
MP
LP

Figure 11:

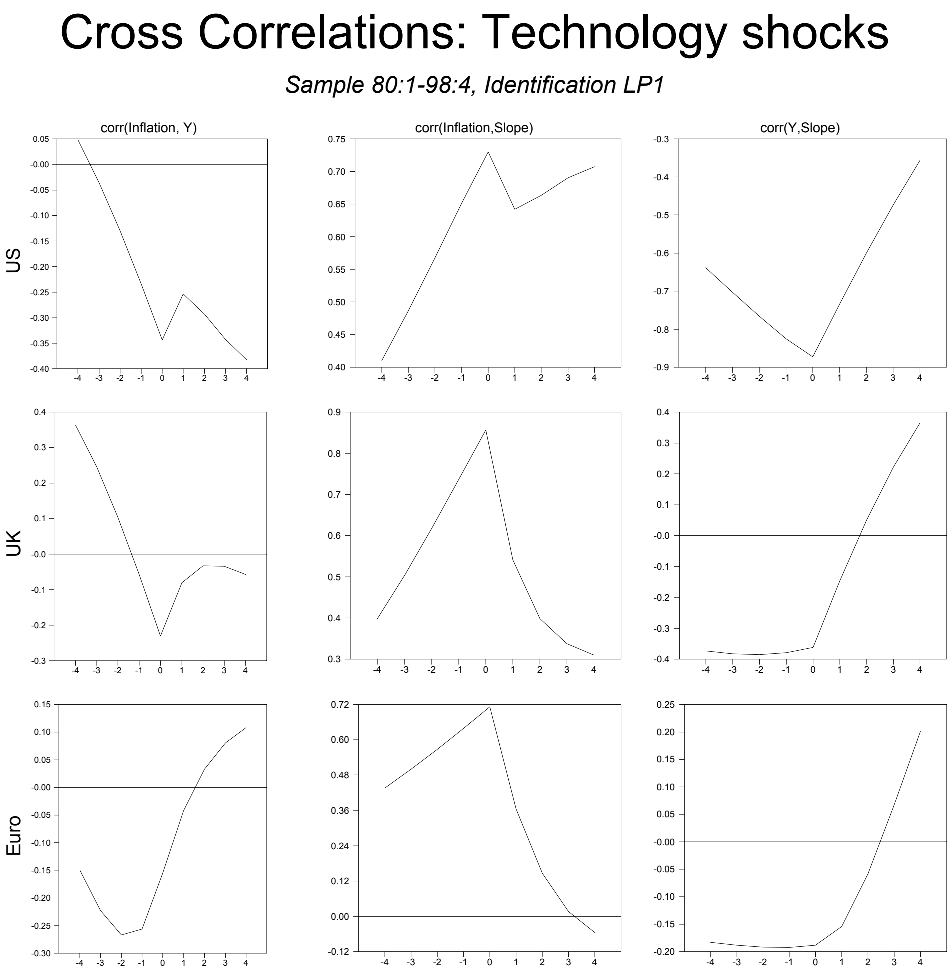

Cross Correlations: Technology shocks
Sample 80:1-98:4, Identification LP1
corr(Inflation, Y)
corr(Inflation,Slope)
corr(Y,Slope)
US
UK
Euro

Figure 12:

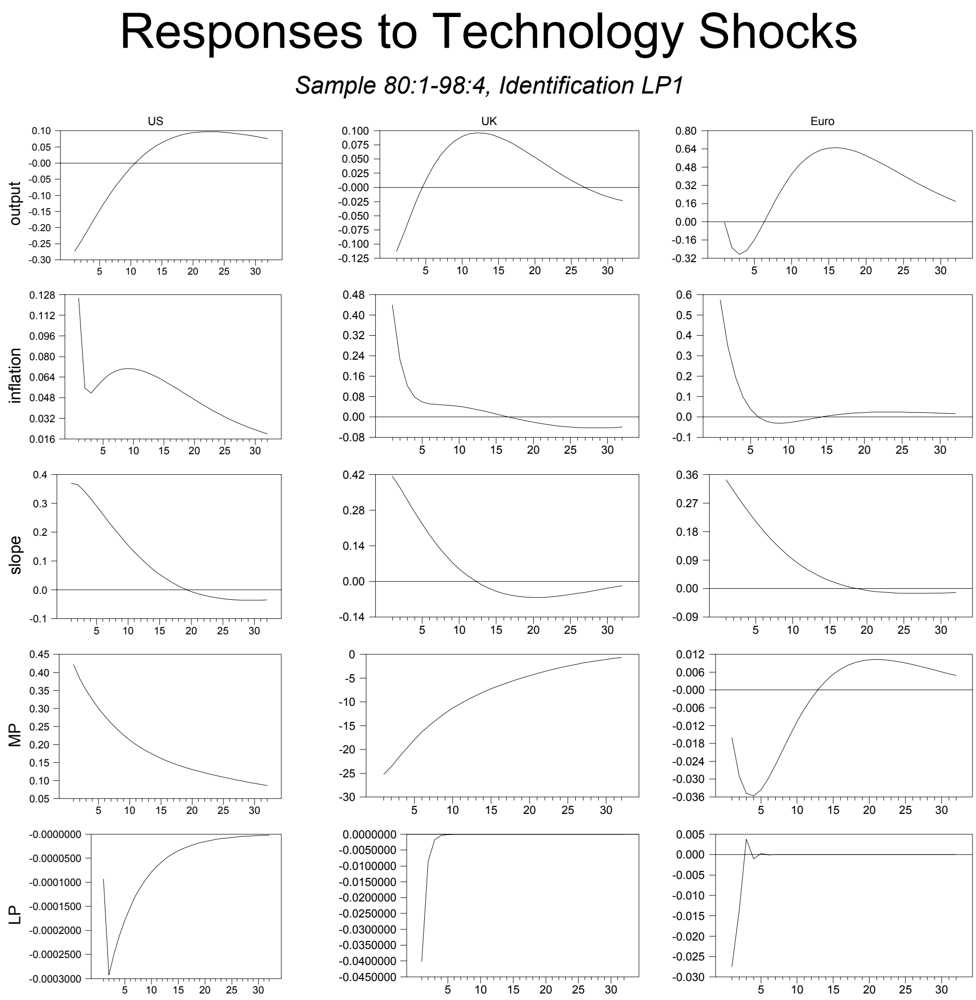

Responses to Technology Shocks
Sample 80:1-98:4, Identification LP1
US
UK
Euro
output
inflation
slope
MP
LP

Figure 13: