Bridging DSGE models and the raw data?

Fabio Canova

May 2012

Barcelona GSE Working Paper Series

Working Paper nº 635

Bridging DSGE models and the raw data

Fabio Canova *
ICREA-UPF, CREI, CREMed and CEPR

May 17, 2012

Abstract

A method to estimate DSGE models using the raw data is proposed. The approach links the observables to the model counterparts via a flexible specification which does not require the model-based component to be solely located at business cycle frequencies, allows the non model-based component to take various time series patterns, and permits model misspecification. Applying standard data transformations induce biases in structural estimates and distortions in the policy conclusions. The proposed approach recovers important model-based features in selected experimental designs. Two widely discussed issues are used to illustrate its practical use.

JEL classification: E3, C3.

Keywords: DSGE models, Filters, Structural estimation, Business cycles.

---

*I would like to thank Bob King (the editor), two anonymous referees, Frank Schorfheide, Tim Cogley, Giorgio Primiceri, Tao Zha, Chris Sims, Harald Uhlig and the participants of seminars at the Bank of England, Bank of Italy, BIS, UCL, Yale, EUI, University of Modena, University of Sussex, Federal Reserve Bank of New York, University of Zurich, ESSIM, the Fed of Atlanta workshop on Methods of applications for DSGE models; the I_t workshop in Time Series Econometrics, Zaragoza; and the conference Recent Development of Dynamic Analysis in Economics, Seoul; The Gent workshop in Macroeconomics for comments and suggestions. The financial support of the Spanish Ministry of Education, through the grant ECO2009-08556, and of the Barcelona Graduate School of Economics is gratefully acknowledged. Department of Economics, UPF, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain, fabio.canova@upf.edu.

1 Introduction

There have been considerable developments in the specification of DSGE models in the last few years. Steps forward have also been made in the estimation of these models. Despite recent efforts, structural estimation of DSGE models is conceptually and practically difficult. For example, classical estimation is asymptotically justified only when the model is the generating process (DGP) of the actual data, up to a set of serially uncorrelated measurement errors, and standard validation exercises are meaningless without such an assumption. Identification problems (see e.g. Canova and Sala, 2009) and numerical difficulties are widespread. Finally, while the majority of the models investigators use is intended to explain only the cyclical portion of observable fluctuations, both permanent and transitory shocks may produce cyclical fluctuations, and macroeconomic data contains many types of fluctuations, and some are hardly cyclical.

The generic mismatch between what models want to explain and what the data contains creates headaches for applied investigators. Over the last 10 years a number of approaches, reflecting different identification assumptions, have been used:

• Fit a model driven by transitory shocks to the observables filtered with an arbitrary statistical device (see Smets and Wouters, 2003, Ireland, 2004a, Rubio and Rabanal, 2005, among others). Such an approach is problematic for at least three reasons. First, since the majority of statistical filters can be represented as a symmetric, two-sided moving average of the raw data, the timing of the information is altered and dynamic responses hard to interpret. Second, while it is typical to filter each real variable separately and to demean nominal variables, there are consistency conditions that must hold - a resource constraint need not be satisfied if each variable is separately filtered - and situations when not all nominal fluctuations are relevant from the point of view of a model. Thus, specification errors can be important. Finally, contamination errors could be present. For example, a Band Pass (BP) filter only roughly captures the power of the spectrum at the frequencies corresponding to cycles with 8-32 quarters average periodicity in small samples and taking growth rates greatly amplifies the high frequency content of the data. In sum, rather than solving the problem, the approach adds to the difficulties applied researchers face.

• Fit a model driven by transitory shocks to transformations of the observables which, in theory, are likely to be void of non-cyclical fluctuations, e.g. consider real "great ratios" (as suggested in Cogley, 2001, and McGrattan, 2010) or nominal " great ratios"(as suggested in Whelan, 2005). As Figure 1 shows, such transformations need not resolve the problem because many ratios still display low frequency movements. In addition, since the number and the nature of the shocks driving non-cyclical fluctuations needs to be a-priori known, specification errors may be produced.

• Construct a model driven by transitory and permanent shocks; scale the model by the assumed permanent shocks; fit the transformed model to the observables transformed in the same way (see e.g. Del Negro et al., 2006, Fernandez and Rubio, 2007, Justiniano, et al., 2010, among others). Such an approach puts stronger faith in the model than previous ones, explicitly imposes consistency between the theory and the observables, but it is not free of problems. For example, since the choice of which shock is permanent is often driven by computational rather than economic considerations, specification errors could be present. In addition, structural parameter estimates may depend on nuisance features, such as the shock which is assumed to be permanent and its time series characteristics. As Cogley (2001) and Gorodnichenko and Ng (2010), have shown, misspecification of these nuisance features may lead to biased estimates of the structural parameters.

• Construct a model driven by transitory and permanent shocks; fit the transformed model to the transformed data in the frequency domain (see e.g. Diebold et. al, 1998, Christiano and Vigfusson, 2003) and select a particular frequency band over which to estimate the structural parameters. This approach is also problematic since it inherits the misspecification problems of the previous approach and the filtering problems of statistically based filtering approaches.

This paper provides an alternative method to estimate DSGE models. I show first that the approach one takes to match the model to the data matters for structural parameter estimation and for economic inference. Unless one has a strong view about what the model is supposed to capture and with what type of shocks, it is difficult to credibly select among various structural estimates (see Canova, 1998). In general, any preliminary data transformations (should these be statistical or model-based) should be avoided if the observed data is assumed to be generated by rational agents maximizing under constraints in a stochastic environment. Statistical filtering does not take into account that the data generated by a DSGE model has power at all frequencies and that, if permanent and transitory shocks are present, the permanent and the transitory component of the data will both appear at business cycle frequencies. Model based transformations impose tight restrictions on the long run properties of the data. Thus, any deviations from the imposed structure, being these residual low frequency variations, unaccounted or idiosyncratic long run dynamics must be captured by the shocks driving the transformed model. Hence, parameter estimates could be distorted because estimates of income and substitution effects could be biased.

The paper proposes to estimate structural parameters by creating a flexible link between the DSGE model and the raw data that allows model based and non-model based components to have power at all frequencies. The methodology can be applied to models featuring transitory or transitory and permanent shocks and only requires that interesting features of the data are left out from the model - these could be low frequency movements of individual series, different long run dynamics of groups of series, etc.. Since the non-model based component can endogenously capture aspects of the data the model is not designed to explain, researchers need not to take a stand on what is left out from the model, or on its time series representation, and therefore shields the analysis from important specification errors. Moreover, because the information present at all frequencies is used in the estimation, filtering distortions are eliminated and inefficiencies minimized. The setup has two other advantages over competitors: structural estimates reflect the uncertainty present in the specification of non-model based features; what the model leaves out at interesting frequencies is easily quantifiable. Thus, R-squared type measures can be built to "test" the structure and to evaluate the explanatory power of additional shocks.

The approach is related to work by Del Negro et al. (2006), in that certain cross equation restrictions that the DGP may impose on the data are not used in estimation, and to the work of Ireland (2004b), in that a non-structural part is added to a structural model prior to estimation and, crucially, it does not substitute for theoretical efforts designed to strengthen the ability of DSGE models to account for all observable fluctuations. But it can fill the gap between what is nowadays available and such a worthy long run aspiration, giving researchers a rigorous tool to address policy questions.

Using a simple experimental design and two practically relevant cases, the paper documents the biases that standard transformations produce, interprets them using the tools developed in Hansen and Sargent (1993), and shows that crucial parameters are better estimated with the proposed procedure. To highlight how the approach can be used in practice, the paper finally examines two questions greatly discussed in macroeconomics: the time variations in the policy activism parameter and the sources of output and inflation fluctuations.

To focus attention on the issues of interest, two simplifying assumptions are made: (i) the estimated DSGE model features no missing variables or omitted shocks and (ii) the number of structural shocks equals the number of endogenous variables. While omitted variables and singularity issues are important in practice, and the semi-structural methods suggested in Canova and Paustian (2011) produce more robust inference when they are present, it is useful to sidestep them because the problems discussed here occur regardless of whether (i)-(ii) are present or not [1].

The rest of the paper is organized as follows. The next section presents estimates of the structural parameters of a simple model when number of statistical and model based

[1] As a referee has pointed out the approach can be used to estimate singular structural models as long as the non-model based component has the same rank as the dimension of the observable variables. Such an extension is not pursued here.

transformations are employed. Section 3 discusses the alternative methodology. Section 4 compares approaches using a simple experimental design. Section 5 examines two economic questions. Section 6 concludes.

2 Estimation with transformed data

To show how estimates of the structural parameters of a DSGE model depend on the preliminary transformation employed, this section considers a textbook small scale New-Keynesian model, where agents face a labor-leisure choice, production is stochastic and requires labour, there is external habit in consumption, an exogenous probability of price adjustments, and monetary policy is conducted with a conventional Taylor rule. Details on the structure are in the on-line appendix.

The model features a technology disturbance z_t, a preference disturbance χ_t, a monetary policy disturbances ϵ_t, and a markup disturbance μ_t. The latter two shocks are assumed to be iid. Depending on the specification z_t, χ are either both transitory, with persistence ρ_z and ρ_χ respectively, or one of them is permanent. The structural parameters to be estimated are: σ_c, the risk aversion coefficient, σ_n, the inverse of the Frisch elasticity, h the coefficient of consumption habit, $1-\alpha$, the share of labor in production, ρ_r, the degree of interest rate smoothing, ρ_π and ρ_y, the parameters of the monetary policy rule, 1-ζ_p, the probability of changing prices. The auxiliary parameters to be estimated are: ρ_χ, ρ_z, the autoregressive parameters of transitory preference and technology shocks, and $\sigma_z, \sigma_\chi, \sigma_r, \sigma_\mu$ the standard deviations of the four structural shocks. The discount factor β and the elasticity among varieties θ are not estimated since they are very weakly identified from the data.

Depending on the properties of the technology and of the preference shocks, the optimality conditions will have a log-linear representation around the steady state or a growth path, driven either by the technology or by the preference shock, see table 1. Four observable variables are used in the estimation. When the model features transitory shocks, parameter estimates are obtained applying four statistical filters (linear detrending (LT), Hodrick and

Prescott filtering (HP), growth rate filtering (FOD) and band pass filtering (BP)) to output, the real wage, the nominal interest rate and inflation. Moreover, three data transformations are employed. In the first, the log of labour productivity, the log of real wages, the nominal rate and the inflation rate, all demeaned, are used as observables (Ratio 1). In the second the log ratio of output to the real wage, the log of hours worked, the nominal rate and the inflation rate, all demeaned, are used as observables (Ratio 2). In the third, the log of the labor share, the log ratio of real wages to output, the nominal interest rate and the inflation rate all demeaned, are used as observables (Ratio 3). When the model features a trending TFP (TFP trend), the linear stochastic specification $z_t = bt + \epsilon_t^z$, is used and the observables for the transformed model are linearly detrended output, linearly detrended wages, demeaned inflation and demeaned interest rates. When the model features trending preferences shocks (Preference trend), the unit root specification, $\chi_t = \chi_{t-1} + \epsilon_t^\chi$ is employed and the observables for the transformed model are the demeaned growth rate of output, demeaned log of real wages, demeaned inflation and demeaned interest rates. Finally, when the model feature a trending TFP, the likelihood function of the transformed model is approximated as in Hansen and Sargent (1993) and only the information at business cycle frequencies $(\frac{\pi}{32}, \frac{\pi}{8})$ is used in the estimation (TFP trend, frequency domain).

The data comes from the FRED database at the Federal Reserve Bank of St. Louis and Bayesian estimation is employed. Since some of the statistical filters are two-sided, a recursive LT filter and a one-sided version of the HP filter have also been considered. The qualitative features of the results are unchanged by this refinement.

Table 2 shows that the posterior distribution of several parameters depend on the preliminary transformation used (see e.g. the risk aversion coefficient σ_c, the Frisch elasticity σ_n^{-1}, the interest smoothing coefficient ρ_r, and persistence and the volatility of the shocks). Since posterior standard deviations are tight, except when estimation is conducted in frequency domain, differences across columns are a-posteriori significant. Posterior differences are also economically relevant. For example, the volatility of markup shocks in the LT, the Ratio

1 and the Preference trend economies is considerably larger and, perhaps unsurprisingly, risk aversion stronger. In addition, when a frequency domain approach is used, the Frisch elasticity is estimated to be very small.

Differences in the location of the posterior of the parameters translate into important differences in the transmission of shocks. As shown in Figure 2, the magnitude of the impact coefficient and of the persistence of the responses vary with the preliminary transformation employed and, for the first few horizons, differences are statistically significant. Furthermore, in the case of technology shocks, the sign of some of the responses is affected.

Why are parameter estimates so different? The first four transformations only approximately isolate business cycle frequencies, leaving measurement errors in the transformed data. In addition, different approaches spread the measurement error across different frequencies: the LT transformation leaves both long and short cycles in the filtered data; the HP transformation leaves high frequencies variability unchanged; the FOD transformation emphasizes high frequency fluctuations and reduces the importance of cycles with business cycle periodicity; and even a BP transformation induces significant small sample approximation errors (see e.g. Canova, 2007). Since the magnitude of the measurement error and its frequency location is transformation dependent, differences in parameter estimates are likely to be important. An approach which can reduce the problematic part of the measurement error is in Canova and Ferroni (2011). More importantly, filtering approaches neglect the fact that the spectral properties of a DSGE model are different from the output of a statistical filter. Data generated by a DSGE model driven by transitory shocks has power at all frequencies of the spectrum and if shocks are persistent most of the power will be in the low frequencies. Thus, concentrating on business cycles frequencies may lead to inefficiencies. Furthermore, when transitory and permanent shocks are present, the transitory and the permanent components of the model will jointly appear in any frequency band and it is not difficult to build examples where, e.g. permanent shocks dominate the variability at business cycle frequencies (see Aguiar and Gopinath, 2007). Hence, the association between the solution of the model and the filtered observables is generally incorrect and biases likely to be generalized.

Implicit or explicit model-based transformations avoid these problems by specifying a permanent and a transitory component of the data with power at all frequencies of the spectrum. However, since specification problems are present (should we use a unit root process or a trend stationary process? Should we allow trending preferences or trending technology?), particular choices lead to nuisance parameters problems (the model estimated with a trending TFP has MA components which do not appear when the preferences are trending, see table 1), and to particular cointegration relationships in the observables, inference depends on the assumptions made and any deviation of the observed data from the assumed structure leads to biases. Finally, frequency domain estimation is inefficient, since most of the variability the model produces is in the low frequencies. Furthermore, while frequency estimation can help to tone down the importance of aspects of the model researchers do not trust, as suggested in Hansen and Sargent (1993), it can not de-emphasize the importance of what the model leaves out at the frequencies of interest.

3 The alternative methodology

Start from the assumption that the observable data has been generated by rational expectation agents, optimizing their objective functions under constraints in a stochastic environment. Assume that the econometrician knows the data generating process for a portion of the data but she is unsure about the transmission produced by certain shocks (e.g. those inducing permanent effects) or how to capture aspects of the data (e.g. those with medium-long period of oscillation). Thus, she is aware that the model used for inference is misspecified. Rather than trying to filter out from the data what the model is unsuited to explain or add ad-hoc features to the model to reduce the misspecification, I will assume that the investigator takes the misspecified structure as given, because it is unclear how to model all the fluctuations present in the data or because the available short cuts are unlikely to satisfactorily account for its complexity. To estimate the parameters of the model she uses the raw data and disregards certain cross equations restrictions present in the DGP but builds a link between the misspecified structural model and the raw data which is sufficiently flexible to capture what the model is unsuited to explain and allows model and non-model based components to jointly appear at all frequencies of the spectrum.

Let the (log)-linearized stationary solution of a DSGE model be of the form:

$$x_{2t} = A(\theta)x_{1t-1} + B(\theta)\epsilon_t \quad (1)$$

$$x_{1t} = C(\theta)x_{1t-1} + D(\theta)\epsilon_t \quad (2)$$

where $A(\theta), B(\theta), C(\theta), D(\theta)$ depend on the structural parameters θ, $x_{1t} \equiv (\log \tilde{x}_{1t} - \log \bar{x}_{1t})$ includes exogenous and endogenous states, $x_{2t} = (\log \tilde{x}_{2t} - \log \bar{x}_{2t})$ all other endogenous variables, ϵ_t the shocks and $\bar{x}_{2t}, \bar{x}_{1t}$ are the long run paths of $\tilde{x}_{2t}$ and $\tilde{x}_{1t}$.

Let $x_t^m(\theta) = R[x_{1t}, x_{2t}]'$ be an $N \times 1$ vector, where R is a selection matrix picking out of x_{1t} and x_{2t} variables which are observable and/or interesting from the point of view of the researcher and let $\bar{x}_t^m(\theta) = R[\bar{x}_{1t}, \bar{x}_{2t}]'$. Let $x_t^d = \log \tilde{x}_t^d - E(\log \tilde{x}_t^d)$ be the log demeaned $N \times 1$ vector of observable data. The specification for the raw data is then:

$$x_t^d = c_t(\theta) + x_t^{nm} + x_t^m(\theta) + u_t \quad (3)$$

where $c_t(\theta) = \log \bar{x}_t^m(\theta) - E(\log \tilde{x}_t^d)$, u_t is a iid $(0, \Sigma_u)$ (proxy) noise, x_t^{nc}, x_t^m and u_t are mutually orthogonal and x_t^{nm} is given by:

$$\begin{aligned} x_t^{nm} &= \rho_1 x_{t-1}^{nm} + w_{t-1} + e_t \qquad e_t \sim \; iid\ (0, \Sigma_e) \\ w_t &= \rho_2 w_{t-1} + v_t \qquad v_t \sim \; iid\ (0, \Sigma_v) \end{aligned} \quad (4)$$

where$\rho_1 = diag(\rho_{11}, ...\rho_{1N}), \rho_2 = diag(\rho_{21}, ...\rho_{2N})$, $0 < \rho_{1i}, \rho_{2i} \leq 1, i = 1, ...N$. To understand what the specification for x_t^{nm} implies, notice that when $\rho_1 = \rho_2 = I$, and e_t, v_t are uncorrelated (4) is the locally linear trend specification used in state space models, see e.g. Gomez (1999). In addition, if $\rho_1 = \rho_2 = I, \Sigma_e$ and Σ_v are diagonal, $\Sigma_{v_i} = 0$, and $\Sigma_{e_i} > 0$, $\forall i$,

x_t^{nm} is a vector of I(1) processes while if $\Sigma_{v_i} = \Sigma_{e_i} = 0$, $\forall i$, x_t^{mn} is deterministic. When instead $\rho_1 = \rho_2 = I$, and Σ_{v_i} and Σ_{e_i} are functions of Σ_ϵ, (4) approximates the double exponential smoothing setup used in discounted least square estimation of state space models, see e.g. Delle Monache and Harvey (2010). Thus, if $\bar{x}_t^m(\theta) = \bar{x}^m(\theta), \forall t$, the observable x_t^d can display any of the typical structures that motivates the use of the statistical filters. Furthermore, as Delle Monache and Harvey (2010) have emphasized, (4) is robust against several types of misspecification of the time series properties of what the model does not explain. Note also, whenever Σ_v is not constrained to be zero, the growth rates of the endogenous variables may display persistent deviations from their mean, a feature that characterizes many real macroeconomic variables, see e.g. Ireland (2010). Finally, when $\bar{x}_t^m(\theta)$ is not constant, and ρ_{1i} and ρ_{2i} are complex conjugates for some i, the specification can capture residual low frequency variations with power at frequency ω. To see this notice that when N=1, (4) implies that $(1-\rho_2 L)(1-\rho_1 L)x_t^{nm} = (1-\rho_2 L)e_t + v_{t-1} \equiv (1-\psi L)\eta_t$. If the roots $\lambda_1^{-1}, \lambda_2^{-1}$of the polynomial $1-(\rho_1+\rho_2)z + \rho_1\rho_2 z^2 = 0$ are complex, they can be written as $\lambda_1^{-1} = r(\cos\omega + i\sin\omega), \lambda_2^{-1} = r(\cos\omega - i\sin\omega)$, where $r = \sqrt{\rho_1\rho_2}$ and $\omega = \cos^{-1}[\frac{\rho_1+\rho_2}{2\sqrt{\rho_1\rho_2}}]$ and (4) is $x_t^{nm} = \sum_j r \frac{sin\omega(j+1)}{sin\omega}(1-\psi L)\eta_t$, whose period of oscillation is $p = \frac{2\pi}{\omega} = \frac{2\pi}{\cos^{-1}[\frac{\rho_1+\rho_2}{2\sqrt{\rho_1\rho_2}}]}$. Thus, given r and p, there exists ρ_1, ρ_2 that produce x_t^{nm} with the required properties.

Given (1)-(4), the data will endogenously select the specification for the non-model based component which is more appropriate for each series and this will be done jointly with the estimation of the structural parameters θ. Identification of the structural parameters is achieved via the cross equation restrictions that the model imposes on the data. Estimates of the non-structural parameters are implicitly obtained from the portion of the data the model can not explain.

The specification has a number of advantages over existing approaches. One does not need to take a stand on the time series properties of the non-model based component and on the choice of filter to tone down its importance and this shields researchers from important specification and filtering errors. As shown in Ferroni (2011), the setup can be used to find the most appropriate specification of the non-model based component and, if a researcher is interested in doing so, to perform Bayesian averaging over different types of non-model based specifications, which is not possible in standard setups. Furthermore, as shown below, all components in (3) may have power at all frequency. Finally, since joint estimation is performed, structural parameter estimates reflect the uncertainty present in the specification of the non-model based component.

3.1 Two special cases

It is useful to consider two special cases of the setup to give a sense of what the approach does. Suppose first that the model features only transitory shocks while the data may display common or idiosyncratic long run drifts, low frequency movements and business cycle fluctuations. Here $\bar{x}^m_t(\theta) = \bar{x}^m(\theta), \forall t$, are the steady states of the model and, if the model is correctly specified on average, $c_t(\theta) = 0$. Assume that no proxy errors are present. Then (3) is

$$x^d_t = x^{nm}_t + x^m_t(\theta) \tag{5}$$

and x^{nm}_t captures the features of x^d_t that the stationary model does not explain. Depending on the specification of ρ_1 and ρ_2, these include long run drifts, both of common and idiosyncratic types, and those idiosyncratic low and business cycle movements the model leaves unexplained. In this setup, x^{nm}_t has two interpretations. As in Altug (1989), McGrattan (1994) and Ireland (2004b), it can be thought of as a measurement error added to the structural model. However, rather than being iid or AR(1), it has the richer representation (4). Alternatively, it can be thought as a reduced form representation for the components of the data the investigator is unsure how to model. Thus, as in Del Negro et al. (2006), x^{nm}_t relaxes the cross equations restrictions that the DGP implies and captures what the model can not explain via the flexible parameterization (4).

Suppose, alternatively, that the model features transitory shocks and one or more permanent shocks. In this case $x^m_t(\theta)$ represents the (stationary) solution in deviation from the permanent shocks and $\bar{x}_t^m(\theta)$ the model based component generated by the permanent shocks. Suppose again that there are no proxy errors. In that case (3) reduces to

$$x_t^d = c_t(\theta) + x_t^{*,nm} + x_t^m(\theta) \tag{6}$$

where $x_t^{*,nm}$ captures the features of x_t^d which neither the transitory portion $x_t^m(\theta)$ nor the permanent portion $c_t(\theta)$ of the model explains. These may include, idiosyncratic long run patterns (such as diverging trends), idiosyncratic low frequency movements, or unaccounted cyclical fluctuations. Comparing (5) and (6), one can see that $x_t^{nm} = c_t(\theta) + x_t^{*,nm}$. Thus, the setup can be used to measure how much of the data the model leaves unexplained and to evaluate whether certain shocks may reduce the discrepancy. For example, one could start from a model featuring a few transitory shocks and measure the relative importance of x_t^{nm} at a particular set of frequencies. If it is large, one could add additional transitory shocks and see how much the relative importance of x_t^{nm} has fallen at those frequencies. Alternatively, one could add a permanent shock and compare the magnitude of $x_t^{*,nm}$ and x_t^{nm} at a particular set of frequencies. By comparing the outcomes of the two exercises, one can also assess whether the addition of a permanent or a transitory shock is more beneficial.

The same logic can be used to evaluate the model when, e.g. the permanent shock takes the form of a stochastic deterministic trend (as in the case of labor augmenting technological progress), when it is represented with a unit root, or when all long run paths are left unmodelled. Hence, the approach naturally provides a setup to judge the goodness of fit of a model and to evaluate the contribution of certain features to the understanding of economic phenomena. It does so by giving researchers a constructive criteria to increase the complexity of models; and an integrated framework to examine the sensitivity of the estimation results to the specification of nuisance features, both of which are absent from existing methods.

3.2 Estimation

Estimation of the structural parameters can be carried out with both classical and Bayesian methods. (1)-(4) can be cast into the linear state space system:

$$s_{t+1} = Fs_t + G\omega_{t+1} \quad \omega_t \sim (0, \Sigma_\omega) \tag{7}$$

$$x_t^d = c_t(\theta) + Hs_t \tag{8}$$

where $s_t = \begin{pmatrix} x_t^{nm} & w_t & x_t^m(\theta) & u_t \end{pmatrix}'$, $\omega_{t+1} = (e_{t+1}, v_{t+1}, u_{t+1}, \epsilon_{t+1})'$, $H = \begin{pmatrix} I & 0 & I & I \end{pmatrix}$, $F = \begin{pmatrix} \rho_1 & I & 0 & 0 \\ 0 & \rho_2 & 0 & 0 \\ 0 & 0 & R[A\ C]' & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $G = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & R[B\ D]' \\ 0 & 0 & I & 0 \end{pmatrix}$. Hence, the likelihood can be computed with a modified Kalman filter (accounting for the possibility of diffuse initial observations) for a given $\vartheta = (\theta, \rho_1, \rho_{2,} \Sigma_e, \Sigma_v, \Sigma_u)$ and maximized using standard tools.

When a Bayesian approach is preferred, one can obtain the non-normalized posterior of ϑ, using standard MCMC tools. For example, the estimates presented in this paper are obtained with a Metropolis algorithm where, given initial ϑ_{-1} and a prior $g(\vartheta)$, candidate draws are obtained from $\vartheta_* = \vartheta_{-1} + \upsilon$, where υ is distributed $t(0, \kappa * \Omega, 5)$ and κ is a tuning parameter, and the draw accepted if the ratio $\frac{\breve{g}(\vartheta_*|y)}{\breve{g}(\vartheta_{-1}|y)}$exceeds a uniform random variable, where $\breve{g}(\vartheta_i|y) = g(\vartheta_i)\mathcal{L}(y|\vartheta_i)$, $i = *, -1$, and $\mathcal{L}(y|\vartheta_i)$ is the likelihood of ϑ_i, . Iterated a large number of times, for κ appropriately chosen, the algorithm ensures that the limiting distribution of ϑ is the target distribution (see e.g. Canova, 2007).

3.3 The relationship with the existing literature

Apart from the work of Del Negro et al. (2006) and of Altug (1989), McGrattan (1994) and Ireland (2004b) already mentioned, the procedure is related to a number of existing works.

First, the state space setup (7)-(8) is similar in spirit to the one suggested by Harvey and Jeager (1993), even thought these authors consider only univariate processes and do not use a structural model to explain the observables. It also shares important similarities with the one employed by Cayen et al. (2009), who are interested in forecasting trends. Two are the most noticeable differences. First, these authors use a two-step estimation approach, conditioning on filtered estimates of the parameters of the DSGE model, while here a one step approach is employed. Second, all the deviations from the model are bundled up in the non-model specification while here it is possible to split them into model interpretable and model non-interpretable parts.

The contribution of the paper is also related to two distinct branches of the macroeconomic and macroeconometric literature. The first attempts to robustify inference when the trend properties of data are misspecified (see Cogley, 2001, and Gorodnichenko and Ng, 2010). I share with the first author the idea that economic theory may not have much to say about certain types of fluctuations but rather than distinguishing between trend stationary and difference stationary cycles, the paper wants to design an estimation procedure which deals with the mismatch between theoretical and empirical concepts of fluctuations without taking a stand on the time series properties of what the model leaves unexplained. The idea of jointly estimating structural and auxiliary parameters without fully specifying the DGP is also present in Gorodnichenko and Ng. However, a likelihood based estimator, as opposed to a minimum distance estimator, is used here because it works regardless of the time series properties of the raw data. In addition, rather than assuming that the model is the DGP, the procedure assumes that the DSGE model is misspecified - a much more useful assumption in practice.

The second branch points out that variations in trend growth are as important as cyclical fluctuations in explaining the dynamics of macroeconomic variables in emerging markets (see e.g. Aguiar and Gopinath, 2007, and Andrle, 2008). While the first paper characterizes differences between emerging and developing economies, the latter is concerned with the misuse of models driven by transitory shocks in policy analyses for developing countries. This paper shows that the problems they highlight are generic and that policy analyses with misspecified models are possible without imposing controversial assumptions about what the model is not designed to explain.

3.4 Setting the priors for Σ_e and Σ_v

If the number of observable variables is small and the number of data points large, one can easily obtain estimates of θ from(7)-(8). If the number of observables is large or the sample size limited, weak identification problems and small sample biases may become relevant. Note, in fact, that in (4) there are $2N + 2N^2$ non-structural parameters to be estimated and that it may be difficult to distinguish variations in the level from variations in the growth rates of the variables. Thus, it may be worth to impose some structure on Σ_v and Σ_e, if information about what the model leaves out is available, and shrewdly cut down on the dimensionality of the non-structural parameter space. For example, one may want to assume that Σ_v and Σ_e are diagonal (so that the non-model based component is series specific), and of reduced rank (the non-model based component is common across (groups of) series); that they have only sparse non-zero elements on the diagonal (the non-model based component exists only in a number of observables) or that they are proportional to each other (shocks to the level and the growth rate are related). Some a-priori restrictions appear to be necessary also because given a DSGE structure, the decomposition of the data in model based and non-model based components depends on the strength of the shock signals. Thus, the procedure defines a family of decompositions, indexed by the relative intensity of the shocks driving the model and the non-model based components. Given that it is typically difficult to estimate this intensity parameter unrestrictedly in small samples, and that unrestricted estimates may imply non-model based components with undesirable high frequency variability, a sensible smoothness prior for Σ_e and Σ_v is needed.

The restrictions which we recommend to be used, and are employed in the two applications described below, involves making Σ_e and Σ_v diagonal, of reduced rank, sparse, and function of the structural shocks. As mentioned, it is possible to approximate the double exponential smoothing restrictions used in discounted least square estimation of state space models by selecting e.g. $\Sigma_{e_i} = \sqrt{\frac{\sigma_\epsilon^2}{\lambda}}$ and $\Sigma_{v_i} = \sqrt{\frac{\sigma_\epsilon^2}{(4\lambda)^2}}$, where i indicates the non-zero elements of the matrices, ϵ_t is one of structural shocks and λ a smoothing parameter. Thus, given a prior for ϵ_t and λ, a prior for all non-zero elements of Σ_e and Σ_v is automatically generated. The specification is attractive because it is parsimonious and considerably reduces the number of non-structural parameters to be estimated. Since λ has the same interpretation as in the HP filter, an agnostic prior for λ could be centered at 400 with uniform range over [4,6400], which allows for very smooth as well as relatively jagged non-model based components [2]. When the likelihood for this parameter is flat, one could alternatively calibrate λ to different values and, in models driven by transitory shocks, eliminate candidates producing non-model based estimates which are not sufficiently smooth. Since one of the structural shocks needs be selected to form the prior for Σ_e and Σ_v, one could also experiment choosing the disturbance with, potentially, the largest or the smallest variance to calibrate the prior. For the applications in section 5, which structural disturbance is employed to calibrate the prior is irrelevant.

In sum, the approach is easy to implement - it requires only a few additional lines in an existing computer code, requires some ingenuity to decrease the dimensionality of the parameter space when the sample is small, but it is otherwise fully operational in practice and, as shown below, it has nice properties in a simple experimental design.

4 The procedure in a controlled experiment

To examine the properties of the procedure and to compare them to those of standard transformations, I use the same setup employed in section 2 and simulate 150 data points assuming that the preference shock has a transitory and a permanent component. Thus, $\chi_t = \chi_{1t} + \chi_{2t}, \chi_{1t} = \rho_\chi \chi_{1t-1} + \epsilon_t^{\chi T}$ and $\chi_{2t} = \chi_{2t-1} + \epsilon_t^{\chi P}$. This specification is chosen since Chang et al. (2007) have indicated that a model with permanent preference shocks can capture well low frequency variations in hours worked. In this setup, the data will display stationary fluctuations driven by four transitory shocks (which we correctly capture with

[2] It is worth noting that selecting the signal to noise ratio λ is much less demanding than assuming a particular format for the drifts the data displays or selecting a shock which drives them.

a model) and important non-stationary fluctuations driven the permanent preference shock (which we will either try to filter out, eliminate with certain data transformations, or account with a non-model based component) making the design relevant for practical purposes. The estimated model is misspecified relative to the DGP in that the permanent component due to the preference shock is left out, but all the other features are correctly represented. Furthermore, since the permanent component of the preference shock is orthogonal to all transitory shocks, the design fits the setup of section 3.

The structural parameters will be estimated using the proposed approach and the same transformations used of section 2 in the most ideal situations one could consider - these include priors centered at the true parameter vector and initial conditions equal to the true parameter vector. When the approach of section 3 is used, the non-model based component is restricted to have a double exponential smoothing format and, consistently with the DGP (see appendix) is allowed to enter only in output and the real wage. The true values of the structural parameters are in table 3. In the estimation the same prior distributions for the structural parameters displayed in table 2 are used. Two cases are examined: one where the permanent disturbance has relative high variability $\sigma_{\chi}^{p} = 1.50$ and one where it has relative low variability $\sigma_{\chi}^{p} = 0.15$. In the first case, the contribution of the permanent component to the spectrum of the series is of the same order of magnitude as the contribution of the transitory component at almost all frequencies. Thus, both filtering and specification errors are present with standard transformations. In the second case, the contribution of the permanent component to the spectrum of the series is everywhere small. Here, standard transformations will only produce filtering errors and, in a large sample, the BP filter provides a consistent although inefficient estimator of model based fluctuations.

As table 3 shows, the distortions produced by standard approaches are important. Apart from producing estimates of utility and technology parameters which are biased and very much filter dependent, the persistence of the preference and of the technology shocks ρ_{χ}, ρ_{z} and the standard deviations of the preference and the markup shocks σ_{χ} and σ_{μ} are generally distorted. In comparison, estimates of utility and technology parameters reported in the column labelled "Flexible" are closer (in a MSE sense) to the true values and both the persistence and the standard deviations of the shocks are better captured. Matching the persistence and the volatility of the shocks is important since conditional and unconditional moments crucially depend on these parameters. Note also that while with standard transformations, estimates depend on the relative intensity of the permanent and transitory signals, this is much less the case for the procedure this paper suggests.

To understand the nature of the distortions produced by standard transformations, note that the log-likelihood of the data can be represented as $L(\theta|y_t) = [A_1(\theta) + A_2(\theta) + A_3(\theta)|y]$, see Hansen and Sargent (1993), where $A_1(\theta) = \frac{1}{\pi}\sum_{\omega_j} \log\det G_\theta(\omega_j)$, $A_2(\theta) = \frac{1}{\pi}\sum_{\omega_j}$ trace $[G_\theta(\omega_j)^{-1}F(\omega_j)]$, $A_3(\theta) = (E(y) - \mu(\theta))G_\theta(\omega_0)^{-1}(E(y) - \mu(\theta))$, $\omega_j = \frac{\pi j}{T}, j = 0, 1, \ldots, T-1$. $G_\theta(\omega_j)$ is the model based spectral density matrix of y_t, $\mu(\theta)$ the model based mean of y_t, $F(\omega_j)$ is the data based spectral density and $E(y)$ the unconditional mean of y_t. $A_2(\theta)$ and $A_3(\theta)$ are penalty functions: $A_2(\theta)$ sums deviations of the model-based from the data-based spectral density over frequencies; $A_3(\theta)$ weights deviations of model-based from data-based means with the spectral density matrix of the model at frequency zero.

Suppose the data is transformed so that the zero frequency is eliminated and the low frequencies de-emphasized. Then, the log-likelihood consists of $A_1(\theta)$ and of $A_2(\theta)^* = \frac{1}{\pi}\sum_{\omega_j}$ trace $[G_\theta(\omega_j)]^{-1}F(\omega_j)^*$, where $F(\omega_j)^* = F(\omega_j)I_{\omega_j}$ and I_{ω_j} is a function describing the effect of the filter at frequency ω_j. Suppose that $I_\omega = I_{[\omega_1,\omega_2]}$, i.e. an indicator function for the business cycle frequencies, as in an ideal BP filter. Then $A_2(\theta)^*$ matters only at business cycle frequencies. Since at these frequencies $[G_\theta(\omega_j)] < F(\omega_j)^*$, $A_2(\theta)^*$ and $A_1(\theta)$ enter additively $L(\theta|y_t)$, two types of biases will be present. Since estimates $\hat{F}(\omega_j)^*$ only approximately capture the features of $F(\omega_j)^*$, $\hat{A}_2(\theta)^*$ has smaller values at business cycle frequencies and a nonzero value at non-business cycle ones. Moreover, in order to reduce the contribution of the penalty function to the log-likelihood, parameters are adjusted so that $[G_\theta(\omega_j)]$ is close to $\hat{F}(\omega_j)^*$ at those frequencies where $\hat{F}(\omega_j)^*$ is not zero. This is done by allowing fitting errors, (a larger $A_1(\theta)$), at frequencies where $\hat{F}(\omega_j)^*$ is zero - in particular, the low frequencies. Hence, the volatility of the structural shocks will be overestimated (this makes $G_\theta(\omega_j)$ close to $\hat{F}(\omega_j)^*$ at the relevant frequencies), in exchange for misspecifying their persistence. These distortions affect agents' decision rules. Higher perceived volatility, for example, implies distortions in the risk aversion coefficient. Inappropriate persistence estimates, on the other hand, imply that perceived substitution and income effects are distorted with the latter typically underestimated. When I_ω is not the indicator function, the derivation of the size and the direction of the distortions is more complicated but the same logic applies. Clearly, different I_ω produce different $\hat{F}(\omega_j)$ and thus different distortions.

Since estimates of $F(\omega_j)^*$ are imprecise, even for large T, there are only two situations when estimation biases are small. First, the permanent component has low power at business cycle frequencies - in this case, the distortions induced by the penalty function are limited. This occurs when transitory volatility dominates (as in the second panel of table 3). Second, when Bayesian estimation is performed, the prior is selected to limit the distortions induced by the penalty function. This is very unlikely, however, since priors are not elicited with such a scope in mind.

If instead one fits a transformed version of the model to transformed data, as it is done in model based approaches, the log-likelihood is composed of $A_1(\theta)^* = \frac{1}{\pi}\sum_{\omega_j} \log |G_\theta(\omega_j) I_{\omega_j}|$ and $A_2(\theta)$ - since the actual and model data are filtered in the same way, the filter does not affect the penalty function. Suppose that $I_\omega = I_{[\omega_1,\omega_2]}$. Then $A_1(\theta)^*$ matters only at business cycle frequencies while the penalty function is present at all frequencies. Therefore, parameter estimates are adjusted so as to reduce the misspecification at all frequencies. Since the penalty function is generally more important at the low frequencies, parameters are selected to make $[G_\theta(\omega_j)]$ close to $\hat{F}(\omega_j)$ at those frequencies and large fitting errors are permitted at medium and high frequencies. Consequently, the volatility of the shocks will be generally underestimated in exchange for overestimating their persistence - somewhat paradoxically, this procedure implies that the low frequency components of the data are those that matter most for estimation. Cross frequency distortions imply that the econometrician recovers an economy which differs substantially from the true one. For example, since less noise is perceived, agents decision rules imply a higher degree of data predictability, and higher perceived persistence implies that perceived substitution and income effects are distorted with the latter overestimated.

To further highlight the properties of the proposed approach, the top row of figure 3 reports estimates of the permanent and transitory components of output obtained with the Kalman filter and either the true parameters or the median estimates presented in the top panel of table 3. The bottom two rows of figure 3 compare the autocorrelation function and the spectral density of the true and estimated components of output.

The true and the estimated components of output display similar volatility properties. In addition, the rate of decay of the autocorrelation functions of the true and the estimated components is practically identical. Finally, as anticipated, the two estimated components have power at all frequencies of the spectrum, and at business cycle frequencies (indicated by the vertical bars in the last row of graphs) the permanent component is more important than the transitory component.

The conditional dynamics in response to transitory shocks are also well captured. Figure 4, which presents impulse responses obtained with true and estimated parameters, indicates that the sign and the persistence of the responses are well matched. Magnitudes are occasionally imprecisely estimated - this problem would remain even if we double the sample size but overall, the approach does a good job in reproducing the main qualitative features of the DGP. Thus, economic inference is less prone to "mismatch" distortions.

5 Two applications

This section shows how the proposed approach can be used to inform researchers about two questions which have received a lot of attention in the literature: the time variations in the policy activism parameter and the sources of output and inflation fluctuations. The first question is analyzed with the model presented in section 2. The second, with a medium scale model widely used in academic and policy circles.

5.1 The policy activism parameter

What are the features of the monetary policy rule in place during the "Great Inflation" of the 1970s and the return to norm of the 1980s and 1990s? This question has been extensively studied in the literature following Clarida et al. (2000). One synthetic way to summarize the information contained in the data is to compute the policy activism parameter $\frac{\rho_y}{\rho_\pi - 1}$, which gives a sense of the relative importance of the output and the inflation stabilization objectives of the Central Bank. The conventional wisdom suggests that the absolute value of this parameter has declined over time, reflecting changes in the preferences of the monetary authorities, but most of the available evidence is obtained either with reduced form methods or, when structural method are used, with filtered data. Are the results to be trusted? Is the characterization offered by the approach of this paper different? Figure 5 plots the posterior density of the policy activism parameter obtained when the data is linearly detrended (top left box) or HP filtered (top right box) before estimation and when the approach of this paper is employed (lower left box) for the samples 1964:1-1979:4 and 1984:1-2007:4. The prior for the structural and auxiliary parameters is the same as in table 1. In the flexible approach, and given the short subsamples, Σ_e and Σ_v are assumed to be diagonal, a common non-model based component is assumed for all the variables, the signal-to-noise ratio in the four series is captured by a single parameter λ, a-priori uniformly distributed over [100, 6400], $\rho_1 = \rho_2 = I$ and the proxy error is set to zero.

The posterior density of the policy activism parameter shifts to the left in the second sample when HP filtered data is used and, for example, the posterior median moves from -0.23 in the first sample to -0.33 in the second. This left shift of the posterior density is absent when LT data is used and the median of the posterior in the second sample moves closer to zero (from -0.38 to 0.12) - care should be exercised here since the median is not a good estimator of the central tendency of the posterior for the 1984-2007 sample. In both cases, the Kolmogorov-Smirnov statistic rejects the null that the posterior distributions are the same in the two samples. Thus, standard approaches confirm the existence of a break in the conduct of monetary policy, although it is not clear in which direction the movement is: with HP filtered data, output gap considerations have become relatively more important; with LT filtered data, the opposite appears to be true.

When the approach of section 3 is used, the posterior density of $\frac{\rho_y}{\rho_\pi - 1}$ in the two samples overlaps considerably. Interestingly, both the location and the shape of the density in the two samples are very similar and the Kolmogorov-Smirnov statistic does not reject the null that the posterior distributions in the two samples are the same. Thus, evidence in favor of a structural break in the conduct of monetary policy is much weaker in this case.

Why are the results different? As mentioned, the non-model based component soaks up all the features that the model is not designed to explain. Thus, in principle, it could absorb the changes present in the endogenous variables. This, however, does not seem to be the case: the median estimate of λ is around 3200 in both samples, making the non-model based component quite smooth relative to the model based component (see on-line appendix for plots of the two components of the four variables) and essentially time invariant. Thus, variations in the time series properties of the endogenous variables are not captured by the non-model based component. What instead happens, is that structural non-policy parameters change to accommodate for the changes in the time series properties of inflation and interest rate over time. Interestingly, the explanatory power of the model increases in the second sub-sample: on average, at business cycle frequencies, the model explains 40 per cent of output variations in the first sample and 55 per cent in the second sample. For inflation and interest rates, the increase is smaller (from 40 to 50 percent).

Since about 50 percent of the variability observables at business cycle frequencies is not captured by the model in both samples, it is worth investigating how the fit can be improved by altering its structure, keeping the number of observables and the estimation approach unchanged. One device that the literature has employed to improve the fit of this kind of models is to allow for a time varying inflation target in the policy rule, see e.g. Ireland (2007). The target is assumed to be driven by a permanent shock and enters only in the interest rate equation. Thus, the estimated specification moves from (5) to (6), where now $c_t(\theta)$ appears only in the interest rate equation. What would this modification do to the posterior distribution of the policy activism parameter?

The last box of figure 5 indicates that adding a time varying inflation target reduces the spread of the posterior distributions. Hence, the shift to the right in the posterior in the second sub-sample becomes statistically significant even though ,e.g., the median value of the two distributions is close in absolute value. Adding an inflation target improves the fit for the interest rate at business cycle frequencies (the proportion of the variance explained increase to 57 percent in the first sample and to 68 percent in the second); for inflation, instead, the explanatory power of the model is unchanged in the first sub-sample and worsen considerably in the second (the variance share explained at business cycle frequencies is now only 28 percent). Hence, adding a time varying inflation target does not seem to be a very promising way to improve our understanding of how inflation fluctuations are generated.

5.2 Sources of output and inflation fluctuations

The question of what drives output and inflation fluctuations has a long history in macroeconomics. In standard medium scale DSGE models, like the one employed by Smets and Wouters (2003) and (2007), output and inflation fluctuations tend to be primarily explained by markup shocks. Since these shocks are an unlikely source of cyclical fluctuations, Chari at al (2009) have argued that misspecification is likely to be present (see Justiniano et al., 2010, for an alternative interpretation). Researchers working in the area use filtering devices to fit the model to the data (as in Smets and Wouters (2003)), arbitrarily data transformations (as in Smets and Wouters, 2007) or build a permanent component in the model (as in Justiniano et al., 2010) and use model-consistent data transformations to estimate the structural parameters. What would the approach of this paper tell us about sources of cyclical fluctuations in output and inflation? How much of the variability of the observables at business cycle frequencies is explained by the model? To answer this question, the same model and the same data set used in Smets and Wouters (2007) are employed but a more standard setup is employed. In particular, no MA terms for the price and wage markup disturbances are assumed - all shocks have a standard AR(1) structure; the model is solved in deviations from the steady state, rather than in deviation from the flexible price equilibrium; and the policy rule does not include a term concerning output growth.

Table 4 reports results obtained eliminating a linear trend from the variables; taking growth rates of the real variables and demeaning nominal ones; and using the approach suggested in this paper. When a linear trend is removed, the forecast error variance decomposition of output at the five years horizon is indeed primarily driven by price markup shocks, with a considerably smaller contribution of investment specific and preference shocks. For inflation, price markup shocks account for almost 90 percent of the forecast error variability at the five years horizon. When the model is instead fitted to growth rates, price markup shocks account for over 90 percent of the variability of both output and inflation at the five years horizon. Thus, even without some of the standard bells and whistles, the conclusion that markup shocks dominate remains. Why are price markup shocks important? Since, compared to other shocks, they are relatively unrestricted in the model, they tend to absorb any misspecification the model has and any measurement error that the filters leave in the transformed data. Furthermore, since the combined specification and measurement errors are unlikely to be iid, the role of markup shocks is overestimated. When the bridge suggested in this paper is used, the non-model based component of real variables is restricted to have a common structure (there are only two parameters simultaneously controlling the non-model based component of output, consumption, investment), $\rho_1 = \rho_2 = I$, and a proxy error is allowed in each equation, the picture is quite different. Output fluctuations at the five year horizon are driven almost entirely by preference disturbances, while inflation fluctuations are jointly accounted for by wage markup, TFP and price markup disturbances. More interestingly, the model explains only 20 percent of the output and inflation fluctuations at business cycle frequencies. Thus, it is seems premature to use it to evaluate policy alternatives.

It is useful to characterize the properties of the non-model based component to evaluate the theoretical modifications that are needed to capture what the current model leaves out. The non-model component is well represented by the specification employed and restrictions on the representation used assuming, for example, no or only one unit root are all rejected in formal testing (log Bayes factor exceeding 10 in both cases). Thus, if shocks are to be added to the model, it is important that they have permanent features and display persistent deviations from a balanced growth path. Ireland (2010) has suggested one such specification. Others, which allow both TFP and investment shocks to have these features, are also possible.

6 Conclusions

Estimating DSGE models with data that is model-based transformed or statistically filtered may lead researchers astray because the association between the output of the filter and the stationary solution of the model is generally incorrect and because model-based transformations impose tight restrictions which are, more likely than not, violated in the data. The consequences of filtering and specification errors could be economically important because income and substitution effects could be distorted, the volatilities and persistence of the shocks over or underestimated and, thus, the decision rules of the agents, as perceived by the econometrician, altered.

The alternative methodology this paper proposes avoids these errors by building a flexible bridge between the DSGE models and the raw data. The procedure is applicable to a large class of models and i) it takes into account the uncertainty in the specification of the non-model component when deriving estimates of the structural parameters; ii) it provides a natural environment to judge the goodness of fit of a model and to evaluate the contribution of certain shocks to the understanding of economic phenomena; iii) it gives researchers an integrated framework to examine the sensitivity of the estimation results to the specification of nuisance features, and iv) it is easy to implement and requires minor modifications of existing routines.

Unaccounted low frequency movements, such as those appearing in hours or labor productivity, or idiosyncratic trends, such as those present in certain relative prices, are typically hard to handle in standard DSGE models. Hence, certain shocks which are left somewhat unrestricted in the model end up capturing these features in standard frameworks. The approach this paper suggests is likely to be very useful in these difficult situations because it helps researchers to distinguish what the model can explain and what it can not, thus avoiding important policy distortions. In general, applications of the methodology appear to be numerous.

Extensions of the setup used in the paper are easy to conceive. For example, structural breaks in the time series features of the observables could be handled either within the model-based (as in Eklund et al., 2008) or the non model-based components and the implications for structural parameters could be compared. Similarly, stochastic volatility could be captured in the model-based or non model-based components and differences evaluated. The unified framework that the approach provides requires very little changes to allow for these situations.

7 Tables and Figures

Model with transitory shocks	
w_t	$= (\frac{\sigma_n}{1-\alpha} + \frac{\sigma_c}{1-h})y_t - \frac{h\sigma_c}{1-h}y_{t-1} - \frac{\sigma_n}{1-\alpha}z_t - \chi_t$
y_t	$= E_t[\frac{1}{1+h}y_{t+1} + \frac{h}{1+h}y_{t-1} - \frac{1-h}{(1+h)\sigma_c}(\chi_{t+1} - \chi_t + r_t - \pi_{t+1})]$
π_t	$= \beta E_t\pi_{t+1} + \frac{1-\alpha}{1-\alpha+\alpha\theta}\frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}(\epsilon_t^\mu + w_t + \frac{\alpha}{1-\alpha}y_t - \frac{1}{1-\alpha}z_t)$
r_t	$= \rho_r r_{t-1} + (1-\rho_r)(\rho_y y_t + \rho_\pi \pi_t) + \epsilon_t^r$
n_t	$= \frac{1}{1-\alpha}(y_t - z_t)$
Model with stochastically trending TFP	
w_t	$= (\frac{\sigma_n}{1-\alpha} + \frac{1}{1-\bar{h}})y_t - \frac{\bar{h}}{1-\bar{h}}y_{t-1} - \chi_t - \frac{\bar{h}}{1-\bar{h}}(\epsilon_{t-1}^z - \epsilon_t^z)$
y_t	$= \frac{1}{1+\bar{h}}E_t(y_{t+1} + hy_{t-1} - (1-\bar{h})(\chi_{t+1} - \chi_t + r_t - \pi_{t+1}) + \bar{h}\epsilon_{t-1}^z + \epsilon_{t+1}^z - (1-\bar{h})\epsilon_t^z)$
π_t	$= \beta E_t\pi_{t+1} + \frac{1-\alpha}{1-\alpha+\alpha\theta}\frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}(\epsilon_t^\mu + w_t + \frac{\alpha}{1-\alpha}y_t)$
r_t	$= \rho_r r_{t-1} + (1-\rho_r)(\rho_y y_t + \rho_\pi \pi_t) + \epsilon_t^r$
n_t	$= \frac{1}{1-\alpha}y_t$
Model with unit roots in preferences	
w_t	$= (\sigma_n + \frac{1}{1-h})y_t - \frac{h}{1-h}y_{t-1} - \sigma_n z_t + \frac{h}{1-h}\epsilon_t^\chi)$
y_t	$=\frac{1}{1+h}E_t(y_{t+1} + hy_{t-1} - (1-h)(r_t - \pi_{t+1}) - (h\epsilon_t^\chi + ((1-h)\sigma_n - h)\epsilon_{t+1}^\chi))$
π_t	$= \beta E_t\pi_{t+1} + \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}(\epsilon_t^\mu + w_t - z_t)$
r_t	$= \rho_r r_{t-1} + (1-\rho_r)(\rho_y y_t + \rho_\pi \pi_t) + \epsilon_t^r$
n_t	$= y_t - z_t$

Table 1: Optimality conditions of the log-linear stationary model. All variables are expressed in percentage deviation from the steady state (balanced growth path). $\bar{h} = e^b h$ and b is the slope of the stochastic trend. With trends $\sigma_c = 1$ and with unit roots in preferences also $\alpha = 0$. z_t is a technology shock, χ_t a preference shock, ϵ_t^r a monetary policy shock and ϵ_t^μ a markup shock. If z_t and χ_t are transitory, $z_t = \rho_z z_{t-1} + \epsilon_t^z, \chi_t = \rho_\chi \chi_{t-1} + \epsilon_t^\chi$. When TFP is trending, $z_t = bt + \epsilon_t^z$, when preferences are trending $\chi_t = \chi_{t-1} + \epsilon_t^\chi$. In each block the first equation defines the equilibrium real wage, the second is an Euler equation, the third a Phillips curve, the fourth a Taylor rule and the fifth a labor demand function.

	Prior	LT	HP	FOD	BP	Ratio 1	Ratio2
		Median (s.e.)	Median (s.e.)	Median (s.e.)	Median (s.e.)	Median(s.e.)	Median (s.e.)
σ_c	$\Gamma(20, 0.1)$	1.90 (0.25)	1.41 (0.21)	0.04 (0.01)	0.96 (0.11)	2.33 (0.27)	0.81 (0.15)
σ_n	$\Gamma(20, 0.1)$	1.75 (0.16)	1.37 (0.13)	5.23 (0.08)	1.19 (0.09)	3.02 (0.24)	2.68 (0.19)
h	$B(6, 8)$	0.83 (0.02)	0.88 (0.02)	0.45 (0.01)	0.96 (0.01)	0.72 (0.05)	0.88 (0.02)
α	$B(3, 8)$	0.07 (0.04)	0.09 (0.05)	0.42 (0.01)	0.07 (0.03)	0.05 (0.04)	0.03 (0.01)
ρ_r	$B(6, 6)$	0.19 (0.05)	0.11 (0.04)	0.62 (0.01)	0.09 (0.02)	0.38 (0.06)	0.28 (0.04)
ρ_π	$N(1.5, 0.1)$	1.33 (0.08)	1.37 (0.05)	1.53 (0.02)	1.51(0.06)	1.92 (0.06)	1.80 (0.05)
ρ_y	$N(0.4, 0.1)$	-0.16 (0.03)	-0.18 (0.03)	0.06 (0.00)	-0.22 (0.03)	0.16 (0.02)	-0.03 (0.02)
ζ_p	$B(6, 6)$	0.82 (0.02)	0.80 (0.03)	0.63 (0.01)	0.86 (0.01)	0.82 (0.02)	0.80 (0.02)
ρ_χ	$B(18, 8)$	0.69 (0.04)	0.40 (0.05)	0.52 (0.01)	0.70(0.02)	0.67 (0.03)	0.66 (0.02)
ρ_z	$B(18, 8)$	0.96 (0.02)	0.95 (0.02)	0.99 (0.01)	0.97(0.01)	0.97 (0.01)	0.96 (0.01)
σ_χ	$\Gamma^{-1}(10, 20)$	0.53 (0.19)	0.47 (0.11)	4.96(0.13)	0.23 (0.05)	3.41 (0.74)	0.97 (0.13)
σ_z	$\Gamma^{-1}(10, 20)$	0.20 (0.04)	0.23 (0.04)	2.00 (0.22)	0.19 (0.03)	0.06 (0.01)	0.06 (0.01)
σ_r	$\Gamma^{-1}(10, 20)$	0.11 (0.01)	0.08 (0.01)	2.30(0.23)	0.07 (0.01)	0.10 (0.01)	0.11 (0.18)
σ_μ	$\Gamma^{-1}(10, 20)$	25.06 (0.97)	14.25 (0.93)	7.17 (0.13)	18.19 (0.66)	22.89 (1.91)	15.94 (0.49)
	Prior	Ratio 3	TFP	Preferences	TFP FD		
		Median (s.e.)	Median (s.e.)	Median (s.e.)	Median (s.e.)		
σ_c	$\Gamma(20, 0.1)$	0.12 (0.03)	1.0	1.0	1.0		
σ_n	$\Gamma(20, 0.1)$	2.09 (0.14)	2.24 (0.26)	2.43 (0.20)	43.17 (23.32)		
h	$B(6, 8)$	0.10 (0.03)	0.08 (0.04)	0.78 (0.03)	0.49 (0.28)		
α	$B(3, 8)$	0.03 (0.02)	0.17 (0.03)	1.0	0.51 (0.28)		
ρ_r	$B(6, 6)$	0.20 (0.06)	0.30 (0.04)	0.61 (0.02)	0.49 (0.28)		
ρ_π	$N(1.5, 0.1)$	1.51 (0.07)	1.74 (0.06)	1.69 (0.05)	1.82 (2.09)		
ρ_y	$N(0.4, 0.1)$	0.77 (0.04)	0.49 (0.03)	0.38 (0.07)	0.09 (2.16)		
ζ_p	$B(6, 6)$	0.81 (0.01)	0.41 (0.03)	0.84 (0.01)	0.48 (0.29)		
ρ_χ	$B(18, 8)$	0.75 (0.03)	0.63 (0.03)		0.48 (0.28)		
ρ_z	$B(18, 8)$	0.62 (0.03)		0.59 (0.02)			
σ_χ	$\Gamma^{-1}(10, 20)$	0.26 (0.04)	0.21 (0.03)	0.06 (0.008)	828.3(81.1)		
σ_z	$\Gamma^{-1}(10, 20)$	0.08 (0.01)	0.05 (0.006)	0.15 (0.02)	284.2 (144.8)		
σ_r	$\Gamma^{-1}(10, 20)$	2.68 (0.27)	0.10 (0.01)	0.07 (0.007)	679.7(232.2)		
σ_μ	$\Gamma^{-1}(10, 20)$	15.98 (1.09)	0.25 (0.04)	36.68 (1.42)	666.9(139.2)		

Table 2: Posterior estimates. LT refers to linearly detrended data, HP to Hodrick and Prescott filtered data, FOD to demeaned growth rates, BP to band pass filtered data. For Ratio 1 the observables are $\log(y_t/n_t), \log(w_t), \pi_t, r_t$, all demeaned, for Ratio 2 they are $\log(y_t/w_t), \log(n_t), \pi_t, r_t$, all demeaned, For Ratio 3, the observables are $\log((w_t n_t)/y_t), \log(w_t/y_t), \pi_t, r_t$, all demeaned. For TFP trending, the observable are linearly detrending output and real wages and demeaned inflation and interest rates. For Preference trending, the observable are demeaned growth rate of output, demeaned log real wages, demeaned inflation and demeaned interest rates. When frequency domain estimation is used, only information in the band $(\frac{\pi}{32}, \frac{\pi}{8})$ is employed. The sample is 1980:1-2007:4.

$\sigma^p_\chi = 1.50$							
	True	LT	HP	FOD	BP	Ratio1	Flexible
		Median (s.e.)	Median (s.e.)	Median (s.e.)	Median(s.e.)	Median(s.e.)	Median(s.e.)
σ_n	0.50	0.12(0.02)	0.21(0.02)	1.30(0.05)	0.08(0.01)	1.00(0.04)	0.24(0.03)
h	0.70	0.91(0.03)	0.74(0.03)	0.71(0.03)	0.88(0.03)	0.11(0.04)	0.76(0.05)
α	0.30	0.07(0.02)	0.06(0.02)	0.04(0.02)	0.16(0.02)	0.04(0.02)	0.20(0.05)
ρ_r	0.70	0.39(0.04)	0.46(0.04)	0.74(0.03)	0.36(0.02)	0.47(0.05)	0.34(0.03)
ρ_π	1.50	1.41(0.06)	1.60(0.06)	1.63(0.06)	1.36(0.05)	1.50(0.08)	1.59(0.08)
ρ_y	0.40	0.01(0.00)	0.01(0.01)	-0.01(0.00)	-0.01(0.00)	0.55(0.07)	-0.01(0.01)
ζ_p	0.75	0.88(0.03)	0.85(0.03)	0.88(0.03)	0.90(0.03)	0.89(0.03)	0.83(0.03)
ρ_χ	0.50	0.40(0.03)	0.36(0.03)	0.69(0.03)	0.73(0.03)	0.37(0.03)	0.51(0.04)
ρ_z	0.80	0.68(0.04)	0.69(0.04)	0.99(0.03)	0.80(0.03)	0.64(0.03)	0.79(0.04)
σ_χ	1.20	3.38(0.41)	0.35(0.06)	0.26(0.05)	0.33(0.12)	0.24(0.04)	0.27(0.07)
σ_z	0.50	0.50(0.11)	0.21(0.04)	0.62(0.11)	0.32(0.06)	0.09(0.01)	0.22(0.04)
σ_r	0.10	0.06(0.01)	0.06(0.01)	0.07(0.01)	0.06(0.01)	0.07(0.01)	0.05(0.00)
σ_μ	1.60	5.97(0.42)	0.80(0.28)	5.60(0.34)	6.62(0.25)	12.33(0.73)	1.56(0.53)
σ_χ	1.20	3.38(0.41)	0.35(0.06)	0.26(0.05)	0.33(0.12)	0.24(0.04)	0.27(0.07)
$\sigma^p_\chi = 0.15$							
	True	LT	HP	FOD	BP	Ratio1	Flexible
		Median (s.e.)	Median (s.e.)	Median (s.e.)	Median(s.e.)	Median(s.e.)	Median(s.e.)
σ_n	0.50	0.18(0.03)	0.35(0.06)	0.89(0.03)	0.31(0.04)	0.95(0.04)	0.14(0.01)
h	0.70	0.92(0.03)	0.91(0.03)	0.90(0.03)	0.97(0.03)	0.13(0.04)	0.79(0.03)
α	0.30	0.05(0.02)	0.07(0.04)	0.23(0.01)	0.14(0.02)	0.03(0.02)	0.15(0.01)
ρ_r	0.70	0.53(0.03)	0.51(0.02)	0.58(0.02)	0.50(0.02)	0.36(0.04)	0.50(0.02)
ρ_π	1.50	1.75(0.07)	1.67(0.06)	1.59(0.05)	1.77(0.06)	1.53(0.07)	1.57(0.05)
ρ_y	0.40	-0.01(0.01)	-0.03(0.01)	-0.03(0.00)	-0.03(0.00)	0.67(0.09)	0.34(0.02)
ζ_p	0.75	0.86(0.03)	0.89(0.03)	0.86(0.03)	0.93(0.03)	0.87(0.03)	0.83(0.03)
ρ_χ	0.50	0.27(0.04)	0.22(0.04)	0.66(0.02)	0.60(0.03)	0.27(0.05)	0.60(0.03)
ρ_z	0.80	0.68(0.04)	0.87(0.03)	0.98(0.03)	0.92(0.03)	0.59(0.05)	0.67(0.03)
σ_χ	1.20	0.39(0.11)	0.31(0.08)	4.23(0.18)	0.30(0.06)	0.18(0.03)	0.85(0.16)
σ_z	0.50	0.23(0.05)	0.22(0.04)	3.37(0.22)	0.17(0.02)	0.06(0.01)	0.22(0.04)
σ_r	0.10	0.06(0.01)	0.06(0.01)	2.61(0.17)	0.06(0.01)	0.07(0.01)	0.07(0.01)
σ_μ	1.60	0.93(0.29)	1.97(0.50)	5.13(0.18)	6.11(0.28)	3.60(0.37)	0.93(0.11)

Table 3: Parameters estimates, simulated data, T=150. LT refers to linearly detrended data, HP to Hodrick and Prescott filtered data, FOD to demeaned growth rates, BP to band pass filtered data, Ratio1 to output scaled by hours, and Flexible to the approach suggested in the paper.

	LT		FOD		Flexible	
	Output	Inflation	Output	Inflation	Output	Inflation
TFP shocks	0.01	0.04	0.00	0.01	0.01	0.21
Gov. expenditure shocks	0.00	0.00	0.00	0.00	0.00	0.02
Investment shocks	0.08	0.00	0.00	0.00	0.00	0.05
Monetary policy shocks	0.01	0.00	0.00	0.00	0.00	0.01
Price markup shocks	0.75(*)	0.88(*)	0.91(*)	0.90(*)	0.00	0.19
Wage markup shocks	0.00	0.01	0.08	0.08	0.03	0.49(*)
Preference shocks	0.11	0.04	0.00	0.00	0.94(*)	0.00

Table 4: Variance decomposition at the 5 years horizon. Estimates are obtained using the median of the posterior of the parameters. A (*) indicates that the 68 percent highest credible set is entirely above 0.10. The model and the data set are the same as in Smets and Wouters, 2007. LT refers to linearly detrended data, FOD to growth rates and Flexible to the approach this paper suggests.

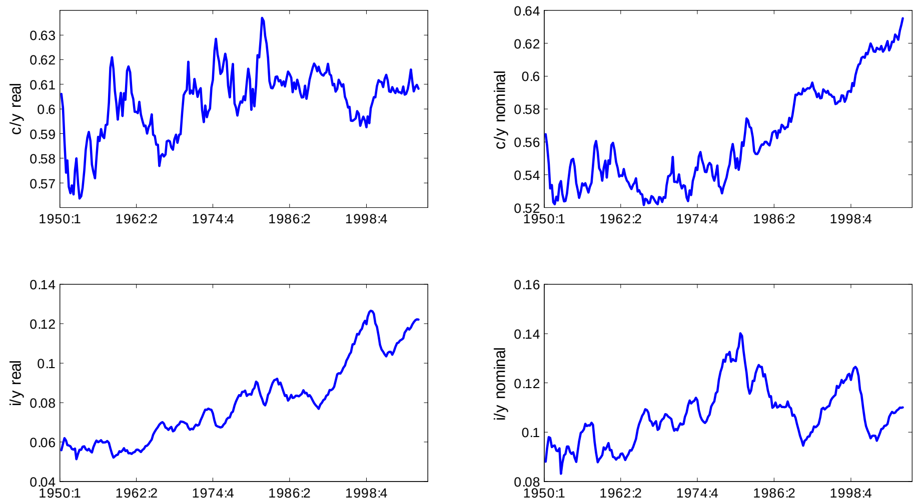

Figure 1: US real and nominal great ratios

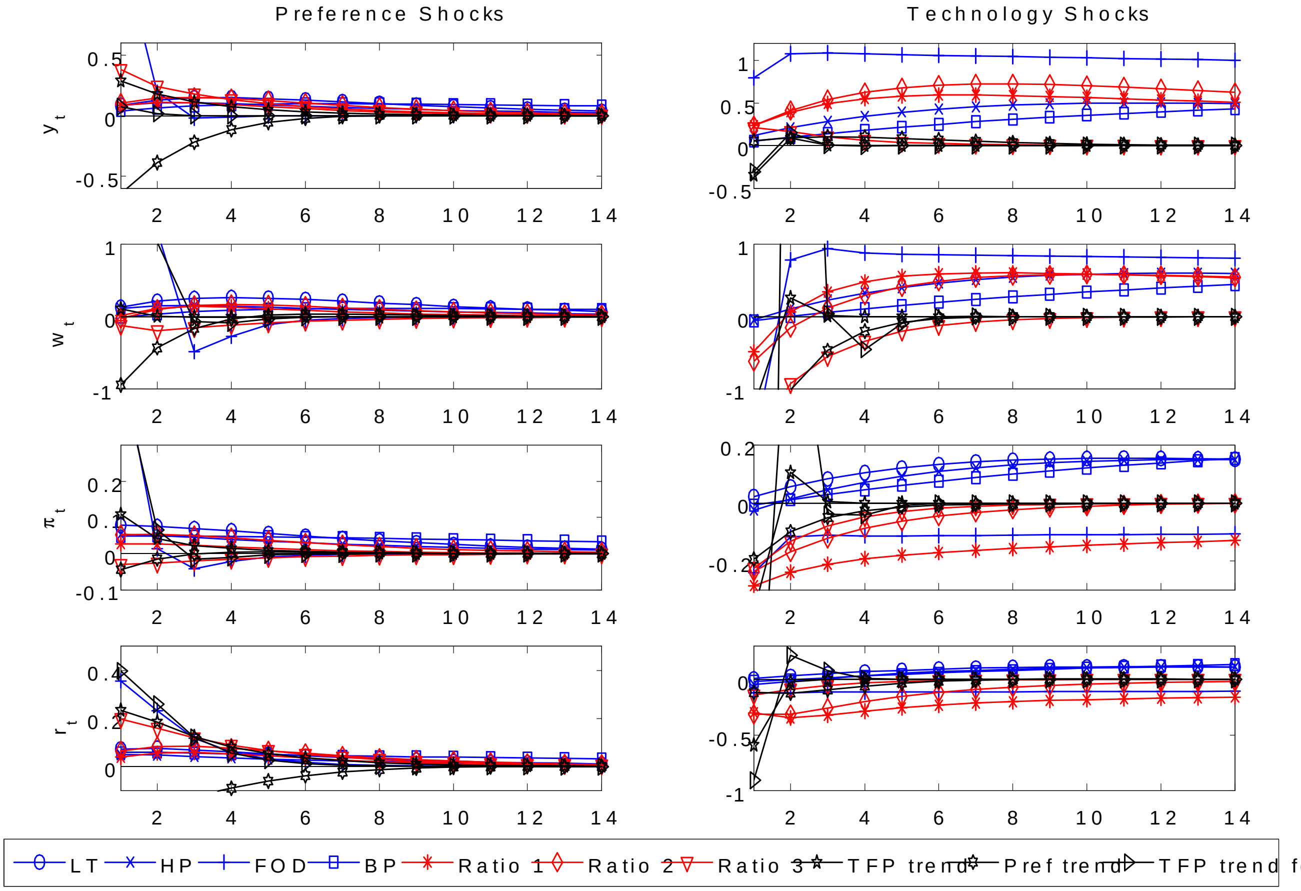

Figure 2: Impulse responses, sample 1980:1-2007:4

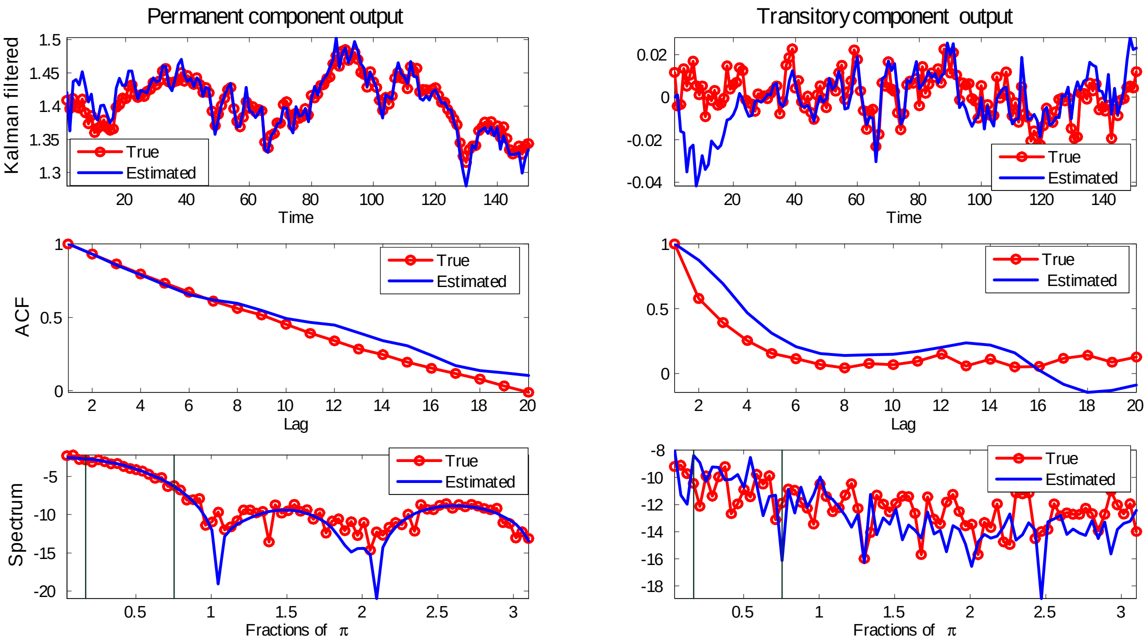

Figure 3: Ouptut decompositions, true and estimated with a flexible approach. Vertical bars indicate cycles with 8-32 quarters periodicity.

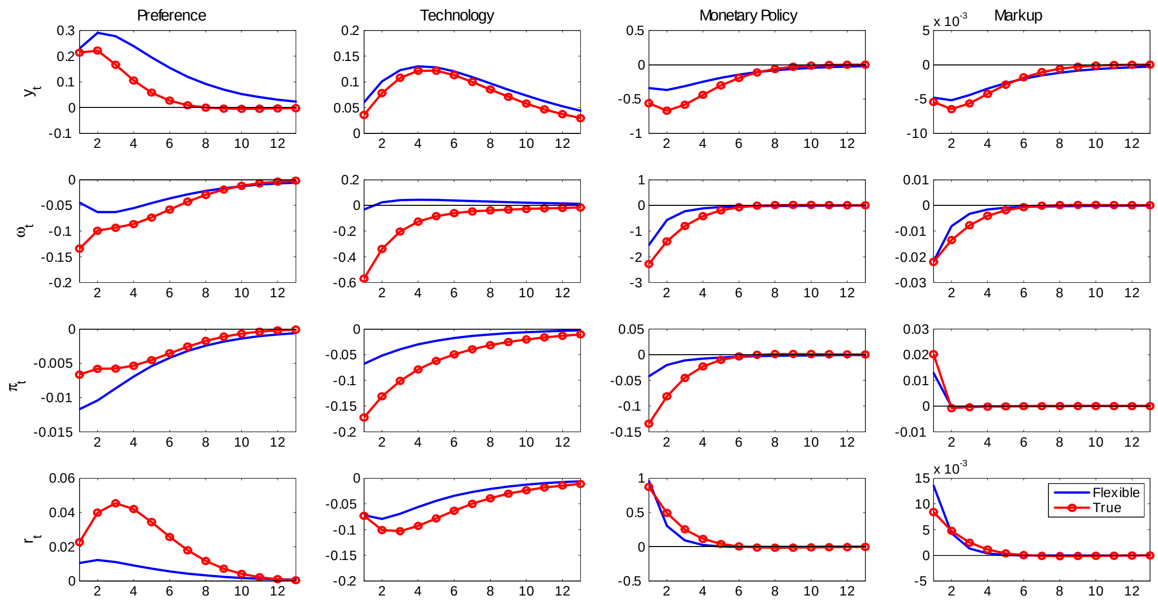

Figure 4: Impulse responses to transitory shocks, true and estimated with flexible approach.

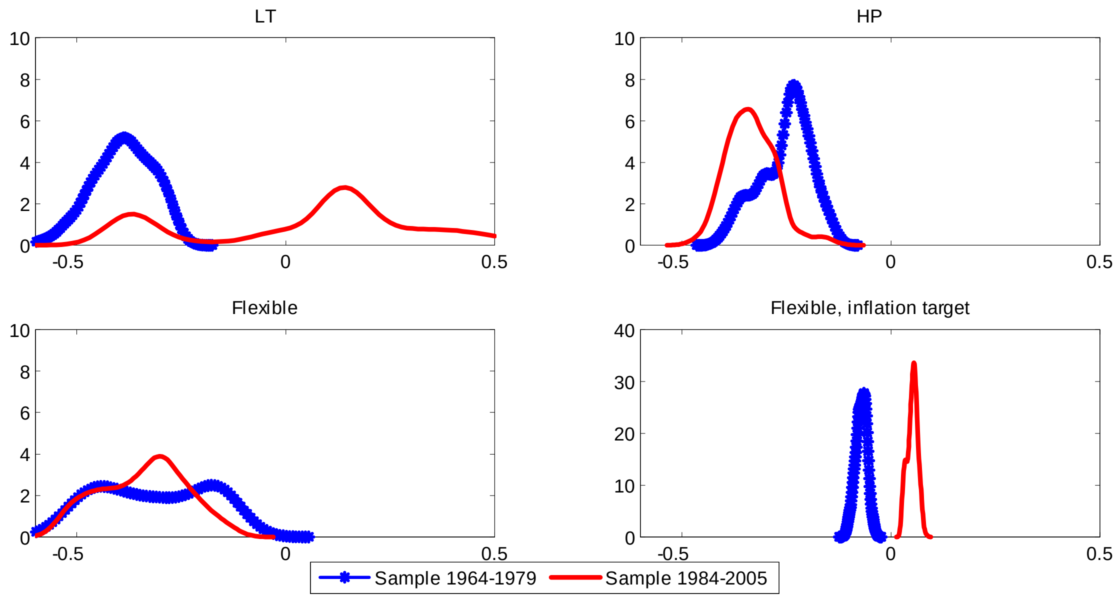

Figure 5: Posterior distributions of the policy activism parameter, samples 1964:1-1979:4 and 1984:1-2007:4. LT refers to linearly detrended data, HP to Hodrick and Prescott filtered data and Flexible to the approach the paper suggests

References

[1] Mark Aguiar and Gita Gopinath, Emerging market business cycles: The cycle is the trend, *Journal of Political Economy*, 115, 69–102, 2007.

[2] Surmu Altug, Time to build and aggregate fluctuations: some new evidence, *International Economic Review*, 20, 883–920, 1989.

[3] Michal Andrle, The role of trends and detrending in DSGE model: Emerging countries need trendy models, 2008.

[4] Fabio Canova, Detrending and business cycles facts, *Journal of Monetary Economics*, 41, 475–512, 1998.

[5] Fabio Canova, *Applied Macroeconomic Research*, Princeton University Press, Princeton, New Jersey, 2007.

[6] Fabio Canova and Filippo Ferroni, Multiple filtering devices for the estimation of cyclical DSGE models, *Quantitative Economics*, 2, 73-98, 2011.

[7] Fabio Canova and Matthias Paustian, Business cycle measurement with some theory, *Journal of Monetary Economics*, 48, 345-361, 2011.

[8] Fabio Canova and Luca Sala, Back to square one: identification issues in DSGE models, *Journal of Monetary Economics*, 56, 431–449, 2009.

[9] Jean Philipppe Cayen, Marc Andre Gosselin and Sharon Kozincki, Estimating DSGE models consistent trends for use in forecasting, Bank of Canada working paper 2009-35, 2009.

[10] Yongsung Chang, Taeyoung Doh, and Frank Schorfheide, Non stationary hours in a DSGE model, *Journal of Money Credit and Banking* , 39, 1357-1373, 2007.

[11] V. V. Chari, Patrick Kehoe and Ellen McGrattan, New Keynesian Models are not yet useful for policy analysis, *American Economic Journal: Macroeconomics*, 1, 242-266, 2009.

[12] Richard Clarida, Jordi Gali and Mark Gertler, Monetary policy rules and Macroeconomic Stability: Evidence and Some Theory *Quarterly Journal of Economics*, 113, 147–180, 2000.

[13] Lawrence J. Christiano and Robert Vigfusson, Maximum likelihood in the frequency domain: the importance of time to plan *Journal of Monetary Economics*, 50, 789–815, 2003.

[14] Timothy Cogley, Estimating and testing rational expectations models when the trend specification is uncertain, *Journal of Economic Dynamics and Control*, 25, 1485–1525, 2001.

[15] Davide Delle Monache and Andrew C. Harvey, The effect of misspecification in models for extracting trends and cycles, manuscript, 2010.

[16] Marco Del Negro, Frank Schorfheide, Frank Smets, and Rafael Wouters, On the fit of New Keynesian models, *Journal of Business and Economic Statistics*, 25, 123–143, 2006.

[17] Frank Diebold, Lee Ohanian and Jeremy Berkovitz, Dynamic equilibrium economies: a framework for comparing models and data, *Review of Economic Studies*, 65, 433-451, 1998.

[18] Jana Eklund, Richard Harrison, George Kapetanios, and Alasdair Scott, Breaks in DSGE models, manuscript, 2008.

[19] Jesus Fernandez Villaverde and Juan Rubio Ramirez, How structural are structural parameter values?, *NBER Macroeconomic Annual*, 24, 83–137, 2008.

[20] Filippo Ferroni, Trend agnostic one step estimation of DSGE models, *B.E. Journal in Macroeconomics*, 11(1), (Advances), Article 25, 2011.

[21] Victor Gomez, Three equivalent methods for filtering finite nonstationary time series, *Journal of Business and Economic Statistics*, 17, 109-116, 1999.

[22] Yuriy Gorodnichenko and Serena Ng, Estimation of DSGE models when the data are persistent, *Journal of Monetary Economics*, 57, 325–340, 2010.

[23] Lars Hansen and Tom Sargent, Seasonality and approximation errors in rational expectations models, *Journal of Econometrics*, 55, 21–55, 1993.

[24] Andrew C. Harvey and Andreas Jaeger, Detrending, stylized facts and the business cycle, *Journal of Applied Econometrics*, 8, 231–247, 1993.

[25] Peter N. Ireland, Technology shocks in the new Keynesian model, *The Review of Economics and Statistics*, 86, 923–936, 2004.

[26] Peter N. Ireland, A method for taking models to the data, *Journal of Economic Dynamics and Control*, 28, 1205–1226, 2004.

[27] Peter N. Ireland, Changes in the Federal Reserve´s Inflation target: causes and consequences, *Journal of Money Credit and Banking*, 38, 1851–1882, 2007

[28] Peter N. Ireland, Stochastic Growth in the US and the Euro Area, forthcoming, *Journal of the European Economic Association*, 2011.

[29] Alejandro Justiniano, Giorgio Primiceri and Andrea Tambalotti, Investment shocks and the Business Cycle, *Journal of Monetary Economics*, 57, 132-145, 2010.

[30] Ellen McGrattan, Measurement with minimal theory, *Federal Reserve Bank of Minneapolis, Quarterly Review*, 2010.

[31] Ellen McGrattan, The macroeconomic effects of distortionary taxation, *Journal of Monetary Economics*, 33, 573-601, 1994.

[32] Pau Rabanal and Juan Rubio, Comparing new Keynesian models of the business cycle: A Bayesian approach, *Journal of Monetary Economics*, 25, 1151–1166, 2005.

[33] Frank Smets and Raf Wouters, An estimated dynamic stochastic general equilibrium model of the euro area, *Journal of European Economic Association*, 1, 1123–1175, 2003.

[34] Frank Smets and Raf Wouters, Shocks and frictions in the US Business Cycle: A Bayesian DSGE approach, *American Economic Review*, 97, 586–606, 2007.

[35] Karl Whelan, New evidence on balanced growth, stochastic trends and economic fluctuations, manuscript, 2005.

On-line Appendix (not intended for publication)

A. The basic DSGE model of section 2

The bundle of goods consumed by the representative household is

$$C_t = \left(\int_0^1 C_t(j)^{\frac{\epsilon_t - 1}{\epsilon_t}} dj \right)^{\frac{\epsilon_t}{\epsilon_t - 1}} \tag{9}$$

where $C_t(j)$ is the consumption of the good produced by firm j and ϵ_t the elasticity of substitution between varieties. Maximization of the consumption bundle, given total expenditure, leads to

$$C_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon_t} C_t \tag{10}$$

where $P_t(j)$ is the price of the good produced by firm j. Consequently, the price deflator is $P_t = \left(\int_0^1 P_t(j)^{1-\epsilon_t} dj \right)^{\frac{1}{1-\epsilon_t}}$ and $P_t C_t = [\int_0^1 P_t(j) C_t(j) dj]$.

The representative household chooses sequences for consumption and leisure to maximize

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[X_t \frac{1}{1-\sigma_c} (C_t - hC_{t-1})^{1-\sigma_c} - \frac{1}{1+\sigma_n} N_t^{1+\sigma_n} \right] \tag{11}$$

where X_t is an exogenous utility shifter following an AR(1) in logs:

$$\chi_t = \rho_\chi \chi_{t-1} + \epsilon_t^\chi \tag{12}$$

where $\chi_t = \ln X_t$ and $\epsilon_t^\chi \sim N(0, \sigma_\chi^2)$. The household budget constraint is

$$P_t C_t + b_t B_t = B_{t-1} + W_t N_t \tag{13}$$

where B_t are one-period bonds with price b_t, W_t is nominal wage and N_t is hours worked.

There is a continuum of firms, indexed by $j \in [0, 1]$, each of which produces a differentiated good. The common technology is:

$$Y_t(j) = Z_t N_t(j)^{1-\alpha} \tag{14}$$

where Z_t is an exogenous productivity disturbance following an AR(1) in log,

$$z_t \quad = \rho_z z_{t-1} + \epsilon_t^z \tag{15}$$

where $z_t = \ln Z_t$ and $\epsilon_t^z \sim N(0, \sigma_z^2)$. Each firm resets its price with probability $1 - \zeta_p$ in any t, independently of time elapsed since the last adjustment. Therefore, aggregate price dynamics are

$$\Pi_t^{1-\epsilon_t} = \zeta_p + (1 - \zeta_p)(P_t^*/P_{t-1})^{1-\epsilon_t} \tag{16}$$

A reoptimizing firm chooses the P_t^* that maximizes the current value of discounted profits

$$\max_{P_t^*} \sum_{k=0}^{\infty} \zeta_p^k E_t Q_{t,t+k} \left[P_t^* Y_{t+k|t} - TC_{t+k}(Y_{t+k|t}) \right] \tag{17}$$

subject to the sequence of demand constraints

$$Y_{t+k|t} = \left(\frac{P_t^*}{P_{t+k}} \right)^{-\epsilon_{t+k}} Y_{t+k} \tag{18}$$

$k = 0, 1, 2, ...$ where $Q_{t,t+k} \equiv \beta^k (C_{t+k}/C_t)(P_t/P_{t+k})$, $TC(.)$ is the total cost function, and $Y_{t+k|t}$ denotes output in period $t + k$ for a firm that reset its price at t.

Finally, the monetary authority sets the nominal interest rate according to

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)(\rho_\pi \pi_t + \rho_y gdp_t) + \epsilon_t^r \tag{19}$$

where $\epsilon_t^r \sim N(0, \sigma_{ms}^2)$.

The first order conditions of the optimization problems are:

$$0 = X_t (C_t - hC_{t-1})^{-\sigma_c} - \lambda_t \tag{20}$$

$$0 = -N_t^{-\sigma_n} - \lambda_t \frac{W_t}{P_t} \tag{21}$$

$$1 = E_t \left[\beta \frac{\lambda_{t+1}}{\lambda_t} \frac{P_{t+1}}{P_t} R_t \right] \tag{22}$$

$$0 = \sum_{k=0}^{\infty} \zeta_p^k E_t Q_{t,t+k} Y_{t+k|t} \left[P_t^* - \mathcal{M}_{t+k} MC_{t+k|t}^n) \right] \tag{23}$$

where λ_t is the Lagrangian multiplier associated with the consumer budget constraint, $R_t \equiv 1 + i_t = 1/b_t$ is the gross nominal rate of return on bonds, $MC^n(.)$ are nominal marginal cost and

$$\mathcal{M}_t = \mu e^{\epsilon_t^\mu} \tag{24}$$

where $\epsilon_t^{\mu} \sim N(0, \sigma_{\mu}^2)$ and μ is the steady state markup.

Market clearing requires

$$\begin{aligned} Y_t(j) &= C_t(j) && (25) \\ N_t &= \int_0^1 N_t(j)dj && (26) \end{aligned}$$

and letting the aggregate output be $GDP_t \equiv \left(\int_0^1 Y_t(j)^{\frac{\epsilon_t - 1}{\epsilon_t}} dj\right)^{\frac{\epsilon_t}{\epsilon_t - 1}}$ we have $C_t = GDP_t$.

The shocks driving the dynamics of the model are: a preference disturbance χ_t, a technology disturbance z_t, a markup shock ϵ_t^{μ} and a monetary shock ϵ_t^r.

B. The solution with transitory shocks

When all the shocks are transitory, the log-linearized equilibrium conditions are:

$$\begin{aligned} w_t &= (\frac{\sigma_n}{1-\alpha} + \frac{\sigma_c}{1-h})y_t - \frac{h\sigma_c}{1-h}y_{t-1} - \frac{\sigma_n}{1-\alpha}z_t - \chi_t && (27) \\ y_t &= E_t[\frac{1}{1+h}y_{t+1} - \frac{h}{1+h}y_{t-1} + \frac{1-h}{(1+h)\sigma_c}(\chi_{t+1} - \chi_t + r_t - \pi_{t+1})] && (28) \\ \pi_t &= \beta E_t \pi_{t+1} + \kappa_p(\epsilon_t^{\mu} + w_t + \frac{\alpha}{1-\alpha}y_t - \frac{1}{1-\alpha}z_t) && (29) \\ r_t &= \rho_r r_{t-1} + (1-\rho_r)(\rho_y y_t + \rho_\pi \pi_t) + \epsilon_t^r && (30) \\ n_t &= \frac{1}{1-\alpha}(y_t - z_t) && (31) \end{aligned}$$

where all variables are expressed in deviation from the (constant) steady state, $k_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}\frac{1-\alpha}{1-\alpha+\psi\alpha}$, $z_t = \rho_z z_{t-1} + \epsilon_t^z$, $\chi_t = \rho_\chi \chi_{t-1} + \epsilon_t^\chi$, ϵ_t^r and ϵ_t^μ are iid. Equation (27) defines the equilibrium real wage, (28) is an Euler equation, (29) a Phillips curve, (30) a Taylor rule and (31) a labor demand function.

This is the model fitted to filtered data (first four columns on the top part of table 2) and to transformed data (the next three columns of table 2).

C. The solution with a stochastic trend in the technology

Assume that the technology has a stochastic linear trend, i.e. $z_t = bt + \epsilon_t^z$, while the other three shocks are assume to be transitory. A log-linearized solution can be found only setting

$\sigma_c = 1$. Defining $\bar{h} = \exp(b)h$, the equations in this case are

$$w_t = \left(\frac{\sigma_n}{1-\alpha} + \frac{1}{1-\bar{h}}\right)y_t - \frac{\bar{h}}{1-\bar{h}}y_{t-1} - \chi_t + \frac{\bar{h}}{1-\bar{h}}(\epsilon_{t-1}^{z,p} - \epsilon_t^{z,p}) \tag{32}$$

$$y_t = \frac{1}{1+\bar{h}}E_t(y_t + hy_{t-1} - (1-\bar{h})(\chi_{t+1} - \chi_t + r_t - \pi_{t+1}) + \bar{h}\epsilon_{t-1}^{z,p} + \epsilon_{t+1}^{z,p} - (1-\bar{h})\epsilon_t^{z,p} \tag{33}$$

$$\pi_t = \beta E_t \pi_{t+1} + \frac{1-\alpha}{1-\alpha-\alpha\theta}\frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}(\epsilon_t^{\mu} + w_t + \frac{\alpha}{1-\alpha}y_t) \tag{34}$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\rho_y y_t + \rho_\pi \pi_t) + \epsilon_t^r \tag{35}$$

$$n_t = \frac{1}{1-\alpha}(y_t - z_t) \tag{36}$$

where all variables are expressed in deviation from the (constant) steady state, $k_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}\frac{1-\alpha}{1-\alpha+\psi\alpha}$, $\chi_t = \rho_\chi \chi_{t-1} + \epsilon_t^\chi$, ϵ_t^r and ϵ_t^μ are iid. Then

$$\ln Y_t - c_y - bt = y_t + \epsilon_t^z \tag{37}$$

$$\ln W_t - c_w - bt = w_t + \epsilon_t^z \tag{38}$$

$$\Pi_t - c_\pi = \pi_t \tag{39}$$

$$R_t - c_r = r_t \tag{40}$$

where capital letters indicate the observable variables, lower case letters the model variables and c_j are constants (the mean of each process). This is the model fitted to the data in column 8 and column 10 of the bottom part of table 2.

D. The solution with non-stationary preference shocks

Assume that $\chi_t = \chi_{t-1} + \epsilon_t^\chi$. A log linearized solution can be found only setting $\sigma_c = 1.0$ and $\alpha = 0$. The log-linearized equilibrium conditions are

$$
\begin{aligned}
w_t &= (\sigma_n + \frac{1}{1-h})y_t - \frac{h}{1-h}y_{t-1} - \sigma_n z_t + \frac{h}{1-h}\epsilon_t^{\chi,p}) && (41)\\
y_t &= \frac{1}{1+h}E_t(y_{t+1} + hy_{t-1} - (1-h)(r_t - \pi_{t+1}) - (h\epsilon_t^{\chi,p} + ((1-h)\sigma_n - h)\epsilon_{t+1}^{\chi,p})) && (42)\\
\pi_t &= \beta E_t \pi_{t+1} + \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}(\epsilon_t^{\mu} + w_t - z_t) && (43)\\
r_t &= \rho_r r_{t-1} + (1-\rho_r)(\rho_y y_t + \rho_\pi \pi_t) + \epsilon_t^r && (44)\\
n_t &= y_t - z_t && (45)
\end{aligned}
$$

where all variables are expressed in deviation from the (constant) steady state, $k_p = \frac{(1-\beta\zeta_p)(1-\zeta_p)}{\zeta_p}$, $z_t = \rho_z z_{t-1} + \epsilon_t^z$, ϵ_t^r and ϵ_t^{μ} are iid. Then

$$
\begin{aligned}
\ln \Delta Y_t - c_y &= y_t + \epsilon_t^{\chi} && (46)\\
\ln W_t - c_w &= w_t && (47)\\
\Pi_t - c_\pi &= \pi_t && (48)\\
R_t - c_r &= r_t && (49)
\end{aligned}
$$

where capital letters indicate the observable variables, lower case letters the model variables and c_j are constants (the mean of the process). This is the model fitted to the data in column 9 of table 2.

E. Simulating data from a model with non-stationary preference shocks

Let Y_t^o be a $N \times 1$ vector of observables and let:

$$
Y_t^o = \nu(\theta^*, \vartheta^*) + H^{ns}x_t^{ns} + H^s x_t^s \qquad (50)
$$

where x_t^s is $N_s \times 1$ vector containing the variables rescaled by the non-stationary preference shock in log deviations from the steady state, $\nu(\theta^*, \vartheta^*)$ is a $N \times 1$ vector of the logarithm of the (rescaled) variables at the steady state, and x_t^{ns} is $N_{ns} \times 1$ vector containing the logarithm of the non-stationary preference shock. H^{ns} is a $N \times N_{ns}$ a selection matrix and H^s is a $N \times N_s$ selection matrix. Finally, $\theta \in \Theta_s$ is the vector of structural parameters describing the stationary dynamics of the DSGE model and $\vartheta \in \Theta_{ns}$ is the vector of parameters that define the non-stationary dynamics. Moreover, $\theta^* \in \Theta_s^* \subset \Theta_s$ and $\vartheta^* \in \Theta_{ns}^* \subset \Theta_{ns}$ are the vectors of parameters that affect the steady state values. Rescaled variables, x_t^s, evolve according to

$$x_{t+1}^s = \Phi(\theta, \vartheta) x_t^s + \Psi(\theta, \vartheta) \eta_{t+1} \qquad \eta_t \sim N(0, \Sigma(\theta, \vartheta)) \tag{51}$$

where η_t is the vector of the structural innovations of the shock processes, $\eta_t = [\eta_t^{ns}, \eta_t^s]'$. It turns out that, for the particular model we have chosen, these equations are given (41)-(45) The vector of non-stationary shock processes $\log X_t^P$is assumed to follows

$$\ln X_t^P = \ln X_{t-1}^P + e_t^{X,P} \tag{52}$$

while the vector of transitory shock processes is

$$\begin{aligned} \log z_t &= \rho_z \log z_{t-1} + e_t^z && (53) \\ \log \chi_t &= \rho_\chi log \chi_{t-1} + e_t^\chi && (54) \\ v_t &= e_t^v && (55) \\ \mu_t &= e_t^\mu && (56) \end{aligned}$$

Thus:

$$x_t^s = [y_t, w_t, \pi_t, r_t, z_t, \chi_t]' \quad (57)$$

$$x_t^{ns} = \ln X_t^P \quad (58)$$

$$\eta_t^s = [e_t^z, e_t^\chi, v_t, \mu_t]' \quad (59)$$

$$\eta_t^{ns} = e_t^{X,P} \quad (60)$$

$$\nu(\theta^*, \vartheta^*) = [\ln y_s, \ln W_s, \ln \Pi_s, \ln R_s]' \quad (61)$$

$$H^{ns} = [1, 1, 0, 0]' \quad (62)$$

$$H^s = \begin{pmatrix} I_{4\times 4} & 0_{4\times 2} \\ & \end{pmatrix} \quad (63)$$

$$\theta = [h, \sigma_n, \rho_r, \rho_y, \rho_\pi, k_p, \rho_z, \rho_\chi, \sigma_z, \sigma_x, \sigma_r, \sigma_\mu] \quad (64)$$

$$\vartheta = \sigma_{X,P} \quad (65)$$

F. The medium scale DSGE model used in section 5

(a): The variables of the model

Label	Definition
y_t	: output
c_t	: consumption
i_t	: investment
q_t	: Tobin's q
k_t^s	: capital services
k_t	: capital
z_t	: capacity utilization
r_t	: real rate
μ_t^p	: price markup
π_t	: inflation rate
μ_t^w	: wage markup
N_t	: total hours
w_t	: real wage rate
R_t	: nominal rate

(b): The parameters of the model

Label	Definition
σ_c	elasticity of intertemporal substitution
σ_l	elasticity of labor supply with respect to real wages
h	habit persistence parameter
δ	depreciation rate
$\phi_p - 1$	share of fixed costs in production
χ	steady state elasticity of capital adjustment cost function
ψ	positive function of the elasticity of capital utilization adjustment costs function.
α	share of capital services in production
γ_p	price indexation parameter
ζ_p	price stickiness parameter
ϵ_p	curvature of good market aggregator
γ_w	wage indexation parameter
ζ_w	wage stickiness parameter
ϵ_w	curvature of labor market aggregator
Label	Definition
λ_r	interest smoothing parameter
λ_π	inflation parameter
λ_y	output parameter
gy	government expenditure to output ratio
ky	steady state capital output ratio
$r_* = \beta^{-1}$	steady state rental rate
w_*	steady state real wage rate
N_*/C_*	steady state hours to consumption ratio

(c): The equations of the model (in deviation from steady states)

$y_t = (1 - gy - \delta\ ky)c_t + \delta\ ky\ i_t + r_*\ ky\ z_t + g_t$	(C.1)
$c_t = \frac{h}{1+h} E_t c_{t+1} + \frac{h}{1+h} c_{t-1} - \frac{(\sigma_c - 1) w_* N_* / C_*}{(1+h)\sigma_c} (N_t - E_t N_{t+1}) - \frac{1-h}{(1+h)\sigma_c} (R_t - E_t \pi_{t+1} + e_t^b)$	(C.2)
$i_t = \frac{\beta}{1+\beta} E_t i_{t+1} + \frac{1}{1+\beta} x_{t-1} + \frac{\chi^{-1}}{1+\beta} q_t + e_t^i$	(C.3)
$q_t = \beta(1-\delta) E_t q_{t+1} + (1 - \beta(1-\delta)) E_t r_{t+1} - (R_t - E_t \pi_{t+1} + e_t^b)$	(C.4)
$y_t = \phi_p(\alpha k_t^s + (1-\alpha) N_t + e_t^a)$	(C.5)
$k_t^s = k_{t-1} + z_t$	(C.6)
$z_t = \frac{1-\psi}{\psi} r_t$	(C.7)
$k_{t+1} = (1-\delta)\ k_t + \delta\ i_t + \delta\ (1+\beta)\ \psi\ e_t^i$	(C.8)
$\mu_t^p = \alpha(k_t^s - N_t) + e_t^a - w_t$	(C.9)
$\pi_t = \frac{\beta}{1+\beta\gamma_p} E_t \pi_{t+1} + \frac{\gamma_p}{1+\beta\gamma_p} \pi_{t-1} - T_p \mu_t^p + e_t^p$	(C.10)
$r_t = -(k_t - N_t) + w_t$	(E.11)
$\mu_t^w = w_t - (\sigma_l N_t + (1-h)^{-1}(c_t - h c_{t-1})$	(C.12)
$w_t = \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta}(E_t \pi_{t+1} + E_t w_{t+1}) - \frac{1+\beta\gamma_w}{1+\beta} \pi_t + \frac{\gamma_w}{1+\beta} \pi_{t-1} - T_w \mu_t^w + e_t^w$	(C.13)
$R_t = \lambda_r R_{t-1} + (1-\lambda_r)(\lambda_\pi \pi_t + \lambda_y y_t) + e_t^r$	(C.14)

The seven disturbances are: TFP shock (e_t^a); monetary policy shock (e_t^r); investment shock (e_t^i); price markup shock (e_t^p); wage markup shock (e_t^w); risk premium shock (e_t^b); government expenditure shock (e_t^g). The compound parameters in equation (C.11) and (C.13) are defined as: $T_p \equiv \frac{1}{1+\gamma_p} \frac{(1-\beta\zeta_p)(1-\zeta_p)}{((\phi_p - 1)\epsilon_p)\zeta_p}$ and $T_w \equiv \frac{1}{1+\beta} \frac{(1-\beta\zeta_w)(1-\zeta_w)}{((\phi_w - 1)\epsilon_w)\zeta_w}$.

(d): The process for the shocks

$e_t = (e_t^a, e_t^r, e_t^i, e_t^p, e_t^w, e_t^b, e_t^g)$
$e_t = \rho e_{t-1} + \eta_t$

where both ρ and $\Sigma = E_t \eta_t \eta_t'$ are diagonal.

G. Additional Tables and Graphs

	LT	HP	FOD	BP
	Median (s.e.)	Median (s.e.)	Median (s.e.)	Median (s.e.)
σ_c	1.68 (0.30)	1.53 (0.26)	0.04 (0.01)	2.98 (0.49)
σ_n	1.73 (0.15)	1.62 (0.12)	5.28 (0.07)	0.55 (0.06)
h	0.85 (0.03)	0.87 (0.03)	0.40 (0.01)	0.89 (0.02)
α	0.05 (0.02)	0.08 (0.03)	0.41 (0.01)	0.04 (0.02)
ρ_r	0.18 (0.06)	0.16 (0.05)	0.64 (0.01)	0.13 (0.03)
ρ_π	1.36 (0.07)	1.36 (0.08)	1.48 (0.02)	1.42 (0.06)
ρ_y	-0.17 (0.03)	-0.17 (0.04)	0.05 (0.00)	-0.11 (0.03)
ζ_p	0.82 (0.01)	0.82 (0.02)	0.64 (0.01)	0.83 (0.01)
ρ_χ	0.66 (0.04)	0.67 (0.04)	0.54 (0.01)	0.81 (0.03)
ρ_z	0.97 (0.02)	0.97 (0.01)	0.99 (0.01)	0.76 (0.02)
σ_χ	0.63 (0.18)	0.65 (0.21)	4.63 (0.07)	0.45 (0.12)
σ_z	0.19 (0.04)	0.23 (0.05)	2.89 (0.19)	0.14 (0.02)
σ_{mp}	0.11 (0.01)	0.11 (0.01)	2.69 (0.14)	0.12 (0.01)
σ_μ	23.13 (1.99)	29.07 (0.94)	7.63 (0.10)	30.22 (1.12)

Table G.1 Parameters estimates obtained with standard transformations; real variables filtered, nominal variables demeaned.

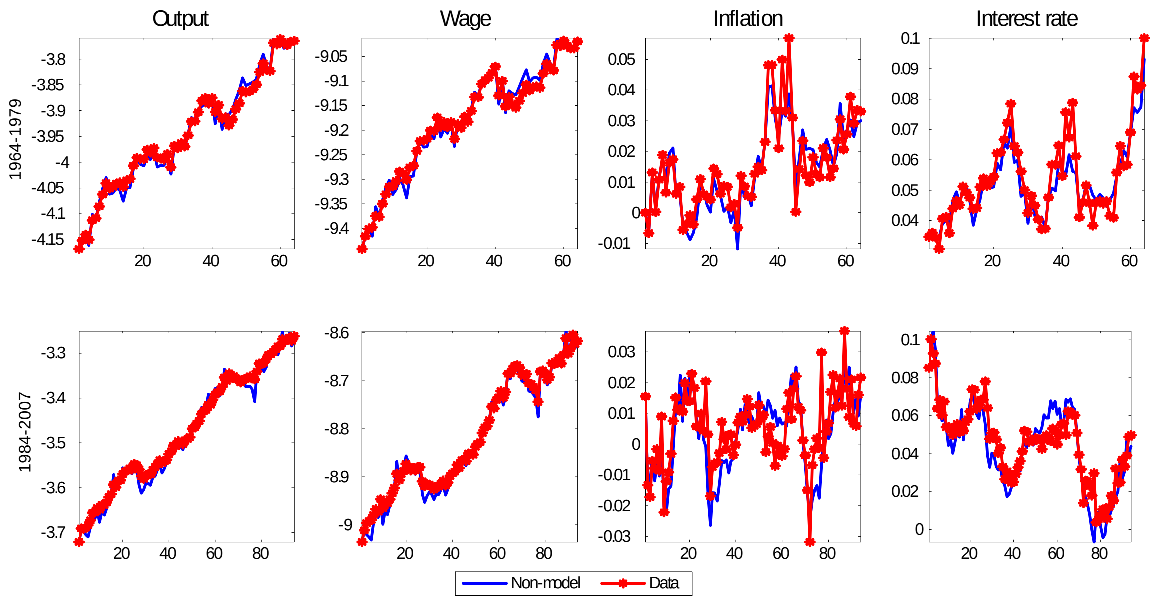

Figure G.2: Data and estimated non-model based components, samples 1964:1-1979:4 and 1984:1-2007:4, flexible approach