WORKING PAPER SERIES

No. 3/2013

Frequentist evaluation of small DSGE models

Gunnar Bårdsen
Department of Economics
Norwegian University of Science and Technology

Luca Fanelli
Department of Statistical Sciences and School of Economics, Statistics and Management,
University of Bologna

Department of Economics

⊡ Norwegian University of Science and Technology
N-7491 Trondheim, Norway
www.svt.ntnu.no/iso/wp/wp.htm

Frequentist evaluation of small DSGE models[*]

Gunnar Bårdsen[†]   Luca Fanelli[‡]

21 February, 2013

Abstract

This paper proposes a new evaluation approach of the class of small-scale 'hybrid' New Keynesian Dynamic Stochastic General Equilibrium (NK-DSGE) models typically used in monetary policy and business cycle analysis. The novelty of our method is that the empirical assessment of the NK-DSGE model is based on a conditional sequence of likelihood-based tests conducted in a Vector Autoregressive (VAR) system in which both the low and high frequency implications of the model are addressed in a coherent framework. The idea is that if the low frequency behaviour of the original time series of the model can be approximated by unit roots, stationarity must be imposed by removing the stochastic trends. This means that with respect to the original variables, the solution of the NK-DSGE model is a VAR that embodies a set of recoverable unit roots/cointegration restrictions, in addition to the cross-equation restrictions implied by the rational expectations hypothesis. The procedure is based on the sequence 'LR1→LR2 →LR3', where LR1 is the cointegration rank test, LR2 the cointegration matrix test and LR3 the cross-equation restrictions test: LR2 is computed conditional on LR1 and LR3 is computed conditional on LR2. The type-I errors of the three tests are set consistently with a pre-fixed overall nominal significance level and the NK-DSGE model is not rejected if no rejection occurs. We investigate the empirical size properties of the proposed testing strategy by a Monte Carlo experiment and illustrate the usefulness of our approach by estimating a monetary business cycle NK-DSGE model using U.S. quarterly data.

Keywords: DSGE models, LR test, Maximum Likelihood, New-Keynesian model, VAR

J.E.L. C5, E4, E5

[†]Department of Economics, Norwegian University of Science and Technology, Dragvoll, 7491, Trondheim.

[‡]Department of Statistical Sciences and School of Economics, Statistics and Management, University of Bologna, via Belle Arti 41, I-40126 Bologna.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models, in the particular the approach dubbed the New Keynesian (NK) models, are presently dominant both in policy and academic environments. It is therefore important to ascertain the validity of these models. Henceforth we denote this class of models with the acronym 'NK-DSGE'.

There are several methods that can be used to evaluate the empirical performance of NK-DSGE models, including economic reliability, statistical fit and forecasting accuracy, see e.g. Schorfheide (2000), An and Schorfheide (2007) and Schorfheide (2011). It is often claimed that Bayesian techniques are preferable to standard likelihood-based methods because NK-DSGE models typically represent a false description of the Data Generating Process (DGP) and misspecification can be important in estimation, see e.g. Canova and Ferroni (2012). Schorfheide (2000) suggests using a loss function to assess the discrepancy between DSGE model predictions and overall posterior distribution of population characteristics that the researcher is trying to match. Del Negro et al. (2007) develop a set of tools within the Bayesian approach that can be used for assessing the time series fit of a DSGE model based on a systematic relaxation of the set of cross-equation restrictions (CER) that the structural model implies on the Vector Autoregressive (VAR) representation of the data.

In this paper we argue that the scientific validity of a model should not be exclusively based on its logical coherence or its intellectual appeal, but also on its capability of making empirical predictions that are not rejected by the data, see e.g. De Grauwe (2010) and Pesaran and Smith (2011). While misspecification in NK-DSGE models is a clear possibility, we do not think it represents a strong argument against the use of frequentist (classical) likelihood-based techniques. A model is purporting by definition to replicate some main properties of its subject, whatever the use or preferences made for its design, and a macroeconometric model should be able to generate the main properties of the actual macro economy.

The purpose of the analysis is therefore to see to what extent it is possible to test the empirical reliability of NK-DSGE models by classical methods. We not only want to use the CER as a 'metric', in line with the spirit of the early literature on the econometrics of rational expectations models, see Hansen and Sargent (1980), Wallis (1980) and Hansen and Sargent (1981), but we also want to test other implied restrictions, usually neglected in the literature, in a coherent and comprehensive framework. Schorfheide (2011) observes that one of the major challenges of current dynamic macro modelling is to recognize that many time series exhibit low frequency behavior that it is difficult to reconcile with the models being estimated. Gorodnichenko and Ng (2010) report in their Table 1 a non-exhaustive listing of how the high persistence of the variables has been addressed in the literature on DSGE models and propose robust estimators

that do not require researchers to take a stand on whether shocks have permanent or transitory effects. We instead account for the low/high frequency behavior of the data by assuming that the variables of interest are driven by unit roots and that the underlying common stochastic trends cancel through steady-state relationships. In this respect, we do not need to take any a priori stand about how to filter the model and the data—because 'filtering' is implicitly obtained by a proper transformation of the model through the cointegration restrictions implied by the theory under investigation.

We use classical frequentist statistical tests, in particular likelihood ratio (LR) tests, with the idea of maximizing the role attached to the data as much as possible. Moreover, we analyze long and short-run restrictions jointly.

We focus on a particular family of small-scale NK-DSGE models typically used in monetary policy and business cycle analysis, see among many others Ireland (2004), Dave and DeJong (2007), Carlstrom et al. (2009), Benati and Surico (2009) and Fanelli (2012) and references therein. We test the NK-DSGE model by a sequential procedure computed in three steps. We start from a finite order VAR involving the (observable) variables of the system. We first test whether the cointegration rank is consistent with the predictions of NK-DSGE model. Next, we test the implied overidentifying cointegrating restrictions, conditional on the chosen rank. Finally, we test the implied overidentifying CER, conditional on the cointegrating restrictions (steady state). Overall, the suggested method involves computing a sequence of three LR tests, hereafter denoted LR1 (LR cointegration rank test), LR2 (LR cointegration matrix test) and LR3 (LR test for CER), leading to a multiple hypothesis testing strategy, whose overall size can be controlled for. For ease of exposition we denote our testing strategy with the symbol 'LR1→LR2→LR3'.

The novelty of the 'LR1→LR2→LR3' approach is that the empirical evaluation of the NK-DSGE model is based on the joint assessment of the low and high frequency implications of the model: LR2 is run conditional upon that LR1 does not reject the cointegration rank and LR3 is run if LR2 does not reject the overidentification cointegrating restrictions.[1] Since we have a precise prediction from the theoretical model about the number of common stochastic trends which should drive the NK-DSGE model, the chosen LR1 test used in the sequence is the 'one

---

[1]To our knowledge, King et al. (1991) and Vredin and Söderlind (1996) are early examples of the use of LR1 in related contexts, Juselius (2011) is a recent example of the use of LR2 in the context of NK-DSGE models, while Guerron-Quintana et al. (2013) propose the inversion of a test like LR3 to build confidence sets for the structural parameters of DSGE models robust to identification failure. Fanelli (2008) applies a testing strategy similar to the one suggested in this paper in a single-equation framework. Fanelli (2012) and Castelnuovo and Fanelli (2011) have recently proposed the use of LR3 (which in the former is actually a Lagrange multiplier test) in the context of NK-DSGE models to test determinacy/indeterminacy.

shot' version of Johansen's LR Trace test, see Johansen (1996).

This paper has several connections with the existing literature. The work closer to ours in spirit is Canova et al. (1994), who propose a method to evaluate real business cycle models by eliciting the (highly) restricted VAR representation underlying them and comparing it with an unrestricted VAR for the data. Canova et al. (1994) recognize that the driving forces in these models may be integrated and hence account for the implied set of cointegration restrictions and also consider what they call 'non-cointegrating restrictions'. Our approach is different from Canova et al. (1994) not only because we focus on a class of monetary policy business cycle models, but mainly because our 'LR1→LR2 →LR3' procedure accounts for long-run and short-run restrictions jointly in a comprehensive framework, allowing us to keep the overall size or the procedure under control. Similarly, Fukač and Pagan (2010) propose an evaluation approach to NK-DSGE models in which both the long and short-run behavior of the data are taken into account by modelling the common stochastic trends in an equilibrium-correction framework. However, while Fukač and Pagan (2010) put forth a 'limited information' approach, our analysis is developed in a 'full information' framework. Juselius (2011) also applies a 'full information' maximum likelihood (ML) approach but he limits attention to the steady-state implications of the NK-DSGE model, leaving the CER untested. Compared to Del Negro et al. (2007), who use a (cointegrated) VAR in error correction form as an approximating model for the DSGE model in their Bayesian evaluation method, we test, other than impose, the cointegration restrictions because testing these restriction is one of the crucial steps of the proposed model evaluation approach. Moreover, Del Negro et al. (2007) consider a model which is a combination of an unrestricted VAR for the data and the VAR subject to the CER implied by the DSGE model: the combination is indexed by a scalar parameter whose level, determined by the data, indicates whether the empirical evidence favours the unrestricted or constrained VAR representation. Instead, the third step of our procedure is explicitly designed to test the CER in the spirit of Hansen and Sargent (1980) and Hansen and Sargent (1981). Compared to Gorodnichenko and Ng (2010), who apply robust filters to both the model and the data, our approach is explicitly designed to testing both the long-run and short-run restrictions and, if the NK-DSGE model is not rejected by the data, delivers ML estimates of the structural parameters. Compared to the likelihood-based estimation and testing approach proposed by Johansen and Swensen (1999) for 'exact' linear rational expectation models (Hansen and Sargent, 1991), the 'LR1→LR2 →LR3' procedure is explicitly focused on the class of monetary policy NK-DSGE models which are prominent examples of 'inexact' linear rational expectations models. 'Inexact' models involve a tighter set of non-linear restrictions compared to their 'exact' counterparts, and this fact complicates the issue of maximizing the constrained likelihood function in just

one step. Moreover, in our set-up the knowledge of the 'right' number of common stochastic trends and the cointegration rank is not taken for granted, but is tested explicitly and is part of the overall testing strategy. In this respect, our approach can also be related to the method put forward by Campbell and Shiller (1987) for estimating and testing present value models through VAR systems. As in Campbell and Shiller (1987), we find the overall set of restrictions that small NK-DSGE models impose on the VAR solution, but differently from Campbell and Shiller (1987) we test all restrictions jointly. Finally, our procedure is very much in the spirit of Hendry and Mizon (1993) who also consider testing of rank, cointegration and overidentifying restrictions in a VAR mapped from non-stationary to stationary representations.

Under the conditions discussed in the paper, the tests LR1, LR2 and LR3, individually considered, are correctly sized in the sense that their asymptotic size is equal to the pre-fixed nominal type I error. Accordingly, using simple Bonferroni arguments, we can easily prove that the overall asymptotic size of the testing strategy does not exceed the sum of the type I errors pre-fixed for each test. Thus, if a practitioner wishes to test the NK-DSGE model at, say, the 5% nominal level of significance, the critical values of the tests LR1, LR2 and LR3 can be chosen such that the sum of the individual type I errors is 5%. The 'LR1→LR2→LR3' test is consistent against all main hypotheses with respect to which its three individual tests are consistent, i.e. (i) DGPs in which the number of common stochastic trends is not consistent with what is implied by the NK-DSGE model; (ii) DGPs in which the number of common stochastic trends is the one predicted by the NK-DSGE model, but the identification structure of the cointegration matrix is at odds with the requirements of the theoretical model; (iii) DGPs in which the CER do not hold, respectively. Undoubtedly, one advantage of the 'LR1→LR2 →LR3' testing strategy is that one can monitor the data adequacy of the NK-DSGE model at the low and high frequency and control at which stage the model is rejected when rejection occurs. The 'LR1→LR2 →LR3' procedure can therefore be regarded as a diagnostic test for the NK-DSGE model by which the researcher can control at which stage the model is rejected and the cause of rejection. Notably, the procedure also delivers, if the NK-DSGE model is not rejected, the ML estimates of the structural parameters.

We discuss the empirical performance of the 'LR1→LR2→LR3' testing strategy by a small Monte Carlo experiment in which the data generating process is assumed to belong to a determinate solution of the structural model of Benati and Surico (2009), which represents the benchmark NK-DSGE specification of our paper. A remarkable by-product of this experiment is the possibility to investigate the identifiability of some of the structural parameters of the NK-DSGE model, in particular those associated with the policy rule, for which an important recent contribution by Cochrane (2011) suggests the impossibility of making reliable inference.

We further show the empirical usefulness of our approach by evaluating the data adequacy of Benati and Surico's (2009) model using U.S. quarterly data. Finally, we report some considerations which might be useful for practitioners.

The paper is organized as follows. We introduce the baseline NK-DSGE model and its assumptions in Section 2 and discuss a set of testable restrictions which are usually ignored in the literature on DSGE models in Section 3. We presents our testing strategy in Section 4 and investigate its empirical size performance by a simulation experiment in Section 5. We present an empirical illustration in which our reference NK-DSGE model is taken to U.S. quarterly data and evaluated empirically in Section 6. A few of suggestions for practitioners are noted in Section 7 that concludes the paper. Two technical appendices summarize some aspects related to the asymptotic properties of the proposed test and the time series representation of the reference NK-DSGE model.

2 Model and assumptions

Our starting point is the structural representation of a typical NK-DSGE model, i.e. the system of equations resulting from the log-linearization around steady-state values of the equations that describe the behavior of economic agents.

Let W_t be the p-dimensional vector collecting all the variables of the model of interest. A typical structural NK-DSGE model which aims at capturing the stylized features of the business cycle takes the form of a linearized rational expectations model:

$$B_0 W_t = B_f E_t W_{t+1} + B_b W_{t-1} + \eta_t^W, \tag{1}$$

where B_0, B_f and B_b are $p \times p$ matrices whose elements depend on the structural parameters collected in the vector θ, and η_t^W is a mean zero vector of disturbances. The term $E_t W_{t+1} = E(W_t \mid \mathcal{F}_t)$ denotes conditional expectations, where $\mathcal{F}_t$ is the available stochastic information set at time t and is such that $\sigma(W_t, W_{t-1}, ..., W_1) \subseteq \mathcal{F}_t$, and $\sigma(W_t, W_{t-1}, ..., W_1)$ is the sigma field generated by the variables.

As is standard in the literature, we posit that η_t^W obeys a vector autoregressive processes of order one, i.e.

$$\eta_t^W = R_W \ \eta_{t-1}^W + u_t^W \ , \quad u_t^W \sim \mathrm{WN}(0_{p \times 1}, \Sigma_{W,u}) \tag{2}$$

where R_W is a stable matrix (i.e. with eigenvalues lying inside the unit disk) and u_t^W is a White Noise disturbance with covariance matrix $\Sigma_{W,u}$. Hereafter u_t^W will be the vector of structural or 'fundamental' disturbances and it will be assumed that $\dim(u_t^W)=\dim(W_t) = $

p, preventing the occurrence of the 'stochastic singularity' issue.[2] Theory does not generally provide indications about the correlation of the structural disturbances across equations; if cross-equation correlations are assumed for the structural disturbances, these can be captured either by specifying a non-diagonal R_W matrix or a non-diagonal $\Sigma_{W,u}$ covariance matrix, or both non-diagonal. We follow the convention of taking R_W to be diagonal, while we allow the possibility of non-diagonal $\Sigma_{W,u}$. The non-zero elements of R_W and of $vech(\Sigma_{W,u})$ enter the vector of structural parameters θ. All meaningful values of θ belong to the 'theoretically admissible' (compact) parameter space, denoted $\mathcal{P}$.

A solution of the model (1)-(2) is any stochastic process $\{W_t^*\}_{t=0}^\infty$, $W_t^* = W_t^*(\theta)$, such that for $\theta \in \mathcal{P}$, $E_t W_{t+1}^* = E(W_{t+1}^* \mid \mathcal{F}_t)$ exists, and for fixed initial conditions, if W_t^* is substituted for W_t into the structural equations, the model is verified for each t. A reduced form solution is a member of the solution set whose time series representation is such that W_t depends on u_t^W, lags of W_t and u_t^W (and, possibly, other arbitrary martingale difference sequences (MDS) with respect to $\mathcal{F}_t$ independent of u_t^W, called 'sunspot shocks').

We confine the class of reduced-form solutions associated with the NK-DSGE model to a known family of linear models by the assumption that follows.

Assumption 1 [Determinacy] The 'true' value θ_0 of θ is an interior point of $\mathcal{P}^*$, where $\mathcal{P}^* \subset \mathcal{P}$ is such that for each $\theta \in \mathcal{P}^*$, the NK-DSGE model (1)-(2) has a determinate reduced-form solution, i.e. unique and asymptotically stationary (stable).

Assumption 1 is crucial to rule out the occurrence of arbitrary parameters unrelated to θ and sunspot shocks unrelated to u_t^W from the time series representation of the system (other than non-stationary explosive processes), see Fanelli (2012). This assumption is standard in the literature on NK-DSGE models and hinges on the idea that the time series upon which model (1) is built and estimated are typically constructed as (or are thought of as being) stationary deviations from steady-state values. In the case of variables such as output, these are mostly log deviations from a steady-state path while, for variables such as interest rates and inflation, they are level deviations from a constant steady-state rate. However, it is well known that removing constants does not ensure stationarity if the persistence of the time series is governed by unit root processes, see Cogley (2001), Juselius and Franchi (2007), Dees et al. (2009), Gorodnichenko and Ng (2010) and Fukač and Pagan (2010). Moreover, treating non-stationary processes mistakenly as stationary may flaw standard inferential procedures, see Johansen (2006), Li (2007) and Fanelli (2008). Thus, for the purpose of testing the model, we will take the implications of

---

[2]One feature of the class of NK-DSGE models considered in this paper is that the number of fundamental shocks is not lower than the number of endogenous variables. In other words, in this setup we do not consider the 'stochastic singularity' issue, see e.g. Ireland (2004) and Dave and DeJong (2007).

Assumption 1 seriously, in the sense that, given the gap between the time series properties observed in the data and Assumption 1, we pursue routes similar to the strategies A and B in Fukač and Pagan (2010), looking for explicit mappings to stationary variables which do not imply loss of information on the low and high frequency behavior of the variables.

We consider a 'fully hybrid' specification of the NK-DSGE system (1)-(2), meaning that in our set-up all diagonal elements of B_b are different from zero. This assumption implies that each Euler equation of the system features at least one lag of the dependent variable.[3]

Under Assumption 1, the unique stable solution of the model (1)-(2) can be represented as the asymptotically stationary VAR system

$$W_t = \tilde{F}_1 W_{t-1} + \tilde{F}_2 W_{t-2} + \varepsilon_t^W \quad , \quad \varepsilon_t^W = \tilde{Q} u_t^W \tag{3}$$

where $\tilde{F}_1 = F_1(\theta)$, $\tilde{F}_2 = F_2(\theta)$ and $\tilde{Q} = Q(\theta)$ are $p \times p$ matrices that depend non-linearly on θ. The dependence is through the implicit set of non-linear CER:

$$(B_0^R - B_f \tilde{F}_1)\tilde{F}_1 - B_f(\tilde{F}_2) + B_{b,1} = 0_{p \times p} \tag{4}$$

$$(B_0^R - B_f \tilde{F}_1)\tilde{F}_2 - B_{b,2} = 0_{p \times p} \tag{5}$$

$$\tilde{\Sigma}_{W,\varepsilon} = \tilde{Q} \, \Sigma_{W,u} \, \tilde{Q}' \tag{6}$$

where $B_0^R = (B_0 + R_W B_f)$, $B_{b,1} = (B_b + R_W B_0)$, $B_{b,2} = -R_W B_b$, $\tilde{Q} = Q(\theta) = \left(B_0 - B_f \tilde{F}_1\right)^{-1}$, and $\tilde{\Sigma}_{W,\varepsilon}$ is the constrained covariance matrix of ε_t^W, see Binder and Pesaran (1995), Uhlig (1999) and Fanelli (2012). We discuss in Appendix B the so-called 'A, B, C's (and D's)' representation associated with determinate reduced form solution in Eq. (3), see e.g. Fernandez-Villaverde et al. (2007) and Ravenna (2007).

The identifiability of the NK-DSGE system depends on whether θ can be uniquely recovered from the mapping in Eq.s (4)-(6), that we compact for ease of exposition in the expression $\phi = g(\theta)$, where $\phi = (vec(F_1)', vec(F_2)', vech(\Sigma_{W,\varepsilon})')'$ is the vector of VAR coefficients and $g(\cdot)$ is a non-linear differentiable function, see Iskrev (2008), Iskrev (2010), and Fanelli (2011). Albeit the function $g(\cdot)$ is not generally available analytically, the Jacobian matrix $\frac{\partial g(\theta)}{\partial \theta'}$ can be evaluated analytically by exploiting the implicit mapping in (4)-(6) and the implicit function theorem, see Iskrev (2008) and Fanelli (2011). If the system information matrix evaluated at θ_0 is non-singular, the ML estimation of θ can be obtained by maximizing the (assumed Gaussian) likelihood of the VAR system (3) subject to numerical approximations of the non-linear constraints in (4)-(6), see among many others, Ruge-Murcia (2007), Dave and DeJong (2007).

---

[3]When $B_b := 0_{p \times p}$ and $R_W := 0_{p \times p}$, system (1)-(2) collapses to a 'purely forward-looking' model. 'Purely forward-looking' models are highly discussed in the monetary policy literature but exhibit problems in the identifiability of some components of θ, see e.g. Lubik and Schorfheide (2003) and Lubik and Schorfheide (2004).

There are cases in which all components in W_t are observed or can be approximated, so the system can be taken directly to the data, see Section 6. In general, however, although the NK-DSGE model in Eq.s (1)-(2) is 'incomplete' as it does not specify how any unobservable components of W_t, denoted $\breve{W}_t$, are generated. Let W_t^o be the sub-vector of W_t that contains the observable variables. Given the n-dimensional 'complete' vector $Z_t = (W_t^{o\prime}, \breve{W}_t')'$ that collects the observable (first) and unobservable variables (last), $n \geq p$, one can interpret the vector W_t in systems (1) and (3) as obtained from the linear combination

$$W_t = \zeta' Z_t \tag{7}$$

where ζ is a known $n \times p$ matrix of full row-rank p that combines the observed and unobserved variables and/or selects the stationary elements of Z_t that enter the structural model. We thus complete the NK-DSGE model by Assumption 2.

Assumption 2 [Unobserved processes are integrated of order one] The vector $\breve{W}_t$ is such that $\Delta \breve{W}_t$ is covariance stationary.

Assumption 2 states that $\breve{W}_t$ is integrated of order one ($\breve{W}_t \sim I(1)$) and can be further specialized as shown in the next sub-section. Given the scope of the present paper, the idea of approximating the unobservable components with $I(1)$ processes meets two requirements. First, in the class of small-scale NK-DSGE models used in monetary policy and business cycle analysis, typical unobservable components are potential output and/or the inflation target (or trend inflation), for which the $I(1)$ hypothesis may represent a reasonable and interpretable choice, as suggested by Bekaert et al. (2010), Fukač and Pagan (2010) and Section 2.1 below. Second, Assumption 2 captures the idea that the persistence observed in many observed time series can be approximated by unit roots processes.

Under Assumptions 1-2, the 'complete' NK-DSGE model is given by

$$A_0 Z_t = A_f E_t Z_{t+1} + A_b Z_{t-1} + \eta_t^Z \tag{8}$$

$$\eta_t^Z = R_Z \eta_{t-1}^Z + u_t^Z \quad , \quad u_t^Z \sim \text{WN}(0_{n \times 1}, \Sigma_{u,Z}), \tag{9}$$

where the matrices A_0, A_f, A_b and $\Sigma_{u,Z}$ now depend on θ and on $\text{diag}(\Sigma_{\breve{W}})$. The 'extended' vector of structural parameters is $\theta^e = (\theta', \theta^{a\prime})'$, where $\theta^a = \text{diag}(\Sigma_{\breve{W}})'$ is the 'additional' $(n-p)$-dimensional vector containing the diagonal elements of $\Sigma_{\breve{W}}$. It is worth emphasizing that the system (8)-(9) also incorporates the model postulated for the unobservable variables, hence it embodies the unit root hypothesis implied by Assumption 2.

The next sub-section provides a detailed example about the relationship between the representation in Eq.s (1)-(2) and (8)-(9) of the NK-DSGE model.

2.1 An example

To focus the discussion, we will throughout use an example based on Benati and Surico (2009). Let $W_t = (\tilde{y}_t, \pi_t, i_t)'$ be the p-dimensional vector ($p = 3$), while the vector of structural shocks is $\eta_t^W = (\eta_{\tilde{y},t}, \eta_{\pi,t}, \eta_{i,t})'$ and the vector of fundamental shocks is $u_t^W = (u_{\tilde{y},t}, u_{\pi,t}, u_{i,t})'$. The model is then made up of the following equations:

$$\tilde{y}_t = \gamma E_t \tilde{y}_{t+1} + (1-\gamma)\tilde{y}_{t-1} - \delta(i_t - E_t \pi_{t+1}) + \eta_{\tilde{y},t} \tag{10}$$

$$\pi_t = \omega_f E_t \pi_{t+1} + \omega_b \pi_{t-1} + \kappa \tilde{y}_t + \eta_{\pi,t} \tag{11}$$

$$i_t = \rho i_{t-1} + (1-\rho)(\varphi_\pi \pi_t + \varphi_y \tilde{y}_t) + \eta_{i,t} \tag{12}$$

$$\eta_{a,t} = \rho_a \eta_{a,t-1} + u_{a,t} \quad , \quad u_{a,t} \sim WN\left(0, \sigma_a^2\right) \quad , \quad a = \tilde{y}, \pi, i \tag{13}$$

In this model, $\tilde{y}_t = (y_t - y_t^p)$ is the output gap, where y_t is the log of output and y_t^p potential output; π_t is the inflation rate and i_t is the nominal interest rate; $\eta_{\tilde{y},t}$, $\eta_{\pi,t}$ and $\eta_{i,t}$ are stochastic disturbances autocorrelated of order one and $u_{\tilde{y},t}$, $u_{\pi,t}$ and $u_{i,t}$ can be interpreted as demand, supply and monetary shocks, respectively. Benati and Surico (2009) restricts the parameters ω_f and ω_b of the New Keynesian Phillips Curve (NKPC) in (11) such that

$$\omega_f = \varrho/(1 + \varrho \varkappa) \quad , \quad \omega_b = \varkappa/(1 + \varrho \varkappa), \tag{14}$$

where ϱ is the firms' discount factor and $\varkappa$ captures the extent of firms' indexation to past prices. This parameterization implies the restriction $\omega_f + \omega_b < 1$. Under these assumptions, the vector of structural parameters is given by $\theta = (\gamma, \delta, \varrho, \varkappa, \rho, \varphi_\pi, \varphi_y, \rho_{\tilde{y}}, \rho_\pi, \rho_i, \sigma_{\tilde{y}}^2, \sigma_\pi^2, \sigma_i^2)'$.

The model (1)-(2) is 'incomplete' as it does not specify how the unobservable component $\breve{W}_t = y_t^p$ is generated. We complete the model by specializing our Assumption 2 as follows.

Assumption 2' [Potential output is a Random Walk]

$$y_t^p = y_{t-1}^p + \eta_{y^p,t} \tag{15}$$

where $\eta_{y^p,t}$ is a white noise term with variance $\sigma_{y^p}^2$.

In addition to the $I(1)$ hypothesis, Assumption 2' captures the need of modelling the unobservable components by simple models whose estimation does not involve many 'extra' parameters to those in θ; in this case, only the variance of potential output $\sigma_{y^p}^2$ is added. The usual interpretation of Assumption 2' is that the flexible price level of output y_t^p is driven by a combination of a stationary demand shock and a non-stationary technology shock, as in Ireland (2004). Moreover, $\theta^a = \mathrm{diag}(\Sigma_{\breve{W}}) = \sigma_{y^p}^2$.

Under Assumption 2', the vector $W_t = (\tilde{y}_t, \pi_t, i_t)'$ featured by the model (10)-(12) can be thought of as being obtained through the linear combination in Eq. (7) which we report here for convenience

$$W_t = \underbrace{\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\zeta'} \underbrace{\begin{pmatrix} y_t \\ \pi_t \\ i_t \\ y_t^p \end{pmatrix}}_{Z_t}, \tag{16}$$

where y_t is output and $Z_t = (W_t^{o\prime}, \breve{W}_t')'$ accommodates both the observable $W_t^o = (y_t, \pi_t, i_t)' = Z_t^o$ and unobserved $\breve{W}_t = y_t^p$ variables. Given the specification of Z_t in Eq. (16), the three equations (10)-(12) jointly with (15) imply the following configuration of the matrices A_0, A_f and A_b of the general model (8)-(9) for the example model

$$A_0 = \begin{pmatrix} 1 & 0 & \delta & -1 \\ -\kappa & 1 & 0 & \kappa \\ -(1-\rho)\varphi_y & -(1-\rho)\varphi_\pi & 1 & (1-\rho)\varphi_y \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A_f = \begin{pmatrix} \gamma & \delta & 0 & -\gamma \\ 0 & \omega_f & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad A_b = \begin{pmatrix} (1-\gamma) & 0 & 0 & -(1-\gamma) \\ 0 & \omega_b & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The 'extended' vector of parameters is defined as $\theta^e = (\theta', \theta^a)'$, where $\theta^a = \sigma_{y^p}^2$.

3   Testable restrictions

Consider the representation in Eq.s (8)-(9) of the NK-DSGE model. Under Assumptions 1 and 2, $Z_t \sim I(1)$. Because of the non-stationarity of Z_t, we need to transform this system such that only stationary variables are involved before its unique stable (determinate) solution can be derived and tested.

To achieve this aim, one possibility is to consider the n-dimensional vector of transformed variables

$$Y_t = \begin{pmatrix} \beta_0' \\ \tau'(1-L) \end{pmatrix} Z_t = G(\beta_0, \tau, 1-L)Z_t, \quad det(\tau'\beta_{0\perp}) \neq 0, \tag{17}$$

where β_0 is the $n \times r$ identified cointegration matrix, and τ is a $(n-r) \times r$ selection matrix which is restricted to be not orthogonal to $\beta_{0\perp}$. The role of τ is to pick out a proper set of variables in

11

first differences from the vector $(1 - L) Z_t = \Delta Z_t$, where L is the lag operator ($L^j Z_t = Z_{t-j}$). The choice of τ in Eq. (17) is not necessarily unique, however. The case discussed below show that despite there are many possible choices of τ, one is consistent with the theoretical features of the NK-DSGE model. In principle, β_0 may temporarily depend on some 'additional' parameters that we collect in the vector ν, and which are not necessarily related to θ. We write $\beta_0 = \beta_0(\nu)$ to make clear such a dependence. Under the null hypothesis that the NK-DSGE model is valid, and with all constraints implied by the NK-DSGE model imposed on the system for Z_t, the joint restriction

$$r = p \ , \ \ \beta_0 = \beta_0^b = \zeta \tag{18}$$

must hold, where the symbol β_0^b denotes the counterpart of the identified cointegration matrix β_0 that leads to what we call a 'balanced' representation of the NK-DSGE model. The condition (18) maintains that under the null hypothesis that the NK-DSGE model is 'true', the identified cointegration matrix β_0 must be equal to the selection matrix ζ introduced in Section 2 and, accordingly, must not depend on any parameter. Hence, the dependence of β_0 on ν is suppressed in Eq.(18). We observe that under the restrictions in Eq. (18) and for a proper choice of the selections matrix τ, the vector Y_t defined in Eq. (17) represents a transformation of the original variables in Z_t which mimics the transformations used by Campbell and Shiller (1987) to address the analysis for present value models.

Under the restriction (18), we can recover W_t from Y_t as follows:

$$Y_t = \begin{pmatrix} \beta_0' \\ \tau'(1 - L) \end{pmatrix} Z_t = G(\beta_0^b, \tau, 1 - L) Z_t$$

$$= \begin{pmatrix} \zeta \\ \tau'(1 - L) \end{pmatrix} Z_t = \begin{pmatrix} \zeta Z_t \\ \tau' \Delta Z_t \end{pmatrix} = \begin{pmatrix} W_t \\ \tau' \Delta Z_t \end{pmatrix}.$$

Hence the vector W_t becomes part of the transformed system for Y_t. Since the $G(\beta_0, \tau, 1 - L)$ (or $G(\beta_0^b, \tau, 1 - L)$) matrix in (17) is non-singular by construction, the representation (17) can be used in the model (8) to obtain

$$A_0 G(\beta_0, \tau, 1 - L)^{-1} Y_t = A_f G(\beta_0, \tau, 1 - L)^{-1} E_t Y_{t+1} + A_b G(\beta_0, \tau, 1 - L)^{-1} Y_{t-1} + \eta_t^Z. \tag{19}$$

The appealing feature of the representation in Eq. (19) is that, other than involving stationary variables (i.e. those in Y_t), the (inverse of the) difference operator $(1 - L)$ cancels out from the equations if one restricts β_0 as in Eq. (18) and imposes a proper set of restrictions on θ such that the transformed model is 'balanced'. With the term 'balanced' we mean that all left-hand and right-hand side variables appearing in system (18) variables are stationary once $G(\beta_0, \tau, 1 - L)^{-1}$ is replaced with $G(\beta_0^b, \tau, 1 - L)^{-1}$ and some restrictions are placed on the elements of θ. The nature of these restrictions will be demonstrated in the two example cases that follow.

12

Hereafter we use the representation

$$A_0^Y Y_t = A_f^Y E_t Y_{t+1} + A_b^Y Y_{t-1} + \eta_t^Y \tag{20}$$

$$\eta_t^Y = R_Y \eta_{t-1}^Y + u_t^Y \tag{21}$$

to denote the 'balanced' counterpart of system (19). The system (20)-(21) can be regarded as an equilibrium-correction representation of the NK-DSGE model and is consistent with the specification strategies A and B in Fukač and Pagan (2010, section 4).

The structural parameters in the matrices A_0^Y, A_f^Y, A_b^Y, R^Y and $\Sigma_{Y,u} = E\left(u_t^Y u_t^{Y\prime}\right)$ are collected in the vector θ^Y, where θ^Y is obtained from θ^e by imposing the restrictions that map system (19) into the transformed representation in (20)-(21). In general, $\dim(\theta^Y) = \dim(\theta^e) - c$, where c is the total number of restrictions on θ^e necessary for balancing.

We now discuss a specific example which helps to clarify the essence of the transformations in Eqs. (17)-(20) and the resulting set of testable restrictions.

Suppose that $r = p = 3$ and that $\beta_0 = \beta_0^b = \zeta$ is specified such that

$$\beta_0^{b\prime} Z_t = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \pi_t \\ i_t \\ y_t^p \end{pmatrix} = W_t. \tag{22}$$

In this case, the output gap, inflation and the short term interest rate are jointly stationary, as typically assumed in small NK-DSGE models. The vector Y_t in Eq. (17) is given by

$$Y_t = G(\beta_0^b, \tau, 1 - L) Z_t = \begin{pmatrix} & \beta_0^{b\prime} & & \\ (1 - L) & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \pi_t \\ i_t \\ y_t^p \end{pmatrix} = (W_t', \Delta y_t)', \tag{23}$$

where it can be noticed that $\tau = (1, 0, 0, 0)'$, $\beta_{0\perp} = (1, 0, 0, 1)'$, and hence $\det(\tau'\beta_{0\perp}) = \det(1) \neq 0$. Correspondingly, the inverse of the transformation matrix $G(\beta_0^b, \tau, 1 - L)$ in Eq. (19) is

$$G(\beta_0^{b\prime}, \tau, 1 - L)^{-1} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{1-L} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & \frac{1}{1-L} \end{pmatrix}.$$

In particular, the four equations of system (19) are given as

$$\tilde{y}_t = \gamma E_t \tilde{y}_{t+1} - \delta(i_t - E_t \pi_{t+1}) + (1-\gamma)\tilde{y}_{t-1} + \eta_{\tilde{y},t}$$

$$\pi_t = \omega_f E_t \pi_{t+1} + \omega_b \pi_{t-1} + \kappa \tilde{y}_t + \eta_{\pi,t}$$

$$i_t = \rho i_{t-1} + (1-\rho)(\varphi_\pi \pi_t + \varphi_y \tilde{y}_t) + \eta_{i,t}$$

$$-\tilde{y}_t + (1-L)^{-1}(1-L)y_t = -\tilde{y}_{t-1} + (1-L)^{-1}(1-L)y_{t-1} + \eta_t^{y^p},$$

where we have left the operator $(1-L)^{-1}$ in the final equation to highlight the point about balancing. To see that $(1-L)^{-1}$ cancels out from this equation, it is sufficient to rewrite it in the form (using Assumption 2)

$$\tilde{y}_t = \tilde{y}_{t-1} + \Delta y_t + \eta_t^{y^p *}$$

where $\eta_t^{y^p *} = -\eta_t^{y^p}$. In this case, $\theta^Y = \theta^e$ and the matrices A_0^Y, A_f^Y and A_b^Y, as well as the vector η_t^Y in the representation in Eq. (20) can easily be derived and are equal to

$$A_0^Y = \begin{pmatrix} 1 & 0 & \delta & 0 \\ -\kappa & 1 & 0 & 0 \\ -(1-\rho)\varphi_y & -(1-\rho)\varphi_\pi & 1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}, \tag{24}$$

$$A_f^Y = \begin{pmatrix} \gamma & -\delta & 0 & 0 \\ 0 & \omega_f & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \quad A_b^Y = \begin{pmatrix} 1-\gamma & 0 & 0 & 0 \\ 0 & \omega_b & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tag{25}$$

and $\eta_t^Y = (\eta_{\tilde{y},t}, \eta_{\pi,t}, \eta_{i,t}, \eta_t^{y^p *})'$. The testable cointegration restrictions relative to the strictly observable time series in Z_t, $Z_t^o = (y_t, \pi_t, i_t)' = W_t^o$, are

$$Z_t^o = (y_t, \pi_t, i_t)' \sim I(1) \text{ and } \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} Z_t^o \sim I(0) \tag{26}$$

or, equivalently, $(\Delta y_t, \pi_t, i_t)' \sim I(0)$.

4 The test sequence

The example discussed in the previous section show one way to obtain the mapping from the complete non-stationary state vector Z_t and the stationary system for Y_t, given the NK-DSGE model and different assumptions about the common stochastic trends driving the variables. It demonstrates that despite the presence of unobservable components, some of the restrictions

14

underlying the mapping from $I(1)$ to $I(0)$ are testable by cointegration analysis, see also Fanelli (2008) and Juselius (2011).

We now turn to our procedure to show how to evaluate the cointegration implications of the NK-DSGE along with other testable implications .Under Assumptions 1-2 (and the other minor assumptions in Section 2), the unique stable solution of the NK-DSGE model (20)-(21) can be represented in the form

$$Y_t = \tilde{P}_1 Y_{t-1} + \tilde{P}_2 Y_{t-2} + \varepsilon_t^Y \quad , \quad \varepsilon_t^Y = \tilde{S} u_t^Y \tag{27}$$

where $\tilde{P}_1 = P_1(\theta^Y)$, $\tilde{P}_2 = P_2(\theta^Y)$ and $\tilde{S} = S(\theta^Y) = (A_0^{Y,R} - A_f^Y \tilde{P}_1)^{-1}$ are $n \times n$ matrices that depend non-linearly on θ^Y through the set of non-linear CER:

$$(A_0^{Y,R} - A_f^Y \tilde{P}_1)\tilde{P}_1 - A_f^Y(\tilde{P}_2) + A_{b,1}^{Y,R} = 0_{n\times n} \tag{28}$$

$$(A_0^{Y,R} - A_f^Y \tilde{P}_1)\tilde{P}_2 - A_{b,2}^{Y,R} = 0_{n\times n} \tag{29}$$

$$\tilde{\Sigma}_{Y,\varepsilon} = \tilde{S} \; \Sigma_{Y,u} \tilde{S}' \tag{30}$$

where $A_0^{Y,R} = (A_0^Y + R_Y A_f^Y)$, $A_{b,1}^{Y,R} = (A_b^Y + R_Y A_0^Y)$, $A_{b,2}^{Y,R} = -R_Y A_b^Y$, and $\tilde{\Sigma}_{Y,\varepsilon}$ is the covariance matrix of the reduced form disturbances ε_t^Y subject to the constraints, see Section 2. The constraints in (28)-(30) mimic those derived in Eq.s (4)-(6) for the 'original' specification of the NK-DSGE model, but refer here to a more general specification in which the role of unobservable components is accounted for by Assumption 2.

Our approach is based on the idea of testing the CER in Eq.s (28)-(30) *without disregarding* the mapping which transforms Z_t into Y_t. As argued in Section 3, the restrictions that lead from Z_t to Y_t are testable by cointegration techniques. This consideration motivates our overall testing strategy.

The null hypothesis is

$$H_0\text{: the DGP belongs to the VAR solution (27)-(30) of the NK-DSGE model} \tag{31}$$

while the alternative hypothesis is

$$H_1\text{: the DGP is not consistent with the VAR solution (27)-(30).} \tag{32}$$

To simplify our exposition without altering the logic of our method, we assume temporarily that all variables in Z_t and Y_t are observable. We turn to the role of unobservables later. Our procedure is based on the following testing steps:

LR1 [Cointegration rank test] We specify a VAR model for Z_t and test for the cointegration rank $r = p$ (corresponding to $n - r$ common stochastic trends driving the system) against

$r = n$ (corresponding to a stationary system), using the 'one shot' version of the LR Trace test (Johansen (1996)). This requires selecting the VAR deterministic components in accordance with the time series features observed in the variables in Z_t. If $r = p$ is rejected, we reject H_0 in Eq. (31) in favour of H_1 in Eq. (32). If instead the selected cointegration rank is found to be equal to the hypothesized rank $r = p$, we consider the next step.

LR2 [Overidentification cointegration restrictions test] Given $r = p$, we fix the (identified) cointegration matrix β_0 at the structure implied by the theoretical model, i.e. $\beta_0 = \beta_0^b = \zeta$, see Eq. (18). Then we compute a LR test for the implied set of overidentifying restrictions, see Johansen (1996). If the LR test rejects the over-identifying restrictions, we reject H_0 in Eq. (31) in favour of H_1 in Eq. (32), otherwise we build the transformed Y_t vector in Eq. (17) by keeping $\beta_0 = \beta_0^b = \zeta$ fixed at the non-rejected structure and consider the next step.

LR3 [Test for CER] We estimate the VAR representation associated with the NK-DSGE model in Eq. (27) by ML both unrestricted (i.e. by leaving P_1, P_2, and $\Sigma_{Y,\varepsilon}$ unrestricted) and subject to the CER in Eq.s (28)-(30). Then we compute a LR test for the CER. If the CER are rejected, we reject H_0 in Eq. (31). If the CER are not rejected, H_0 is accepted and $\hat{\theta}^Y$ is the ML estimator of the structural parameters of the NK-DSGE model.

Albeit to some extent cointegrated VAR models have been already used to evaluate DSGE models through 'frequentist' approaches, see e.g. Canova et al. (1994), the 'LR1→LR2 →LR3' sequence is a novel approach in the literature.[4] The null H_0 in Eq. (31) can be rejected either because the model predicted cointegration rank is rejected (LR1), or because the predicted model structure of cointegration relationships is rejected (LR2) given that the cointegration rank implied by the model is accepted, or because the CER are rejected (LR3) given that the predicted model structure of cointegration relationships is not rejected. The null hypothesis is not rejected if all three tests pass.

---

[4]Rational expectation models are often given an 'exact' representation, i.e. such that the structural disturbances η_t^W in (1) are absent, see Hansen and Sargent (1991). Johansen and Swensen (1999) have shown how 'exact' linear rational expectations models, which feature non-stationary observable variables, can be nested within the cointegrated VAR model and tested by likelihood-based methods in just one solution. Given the 'exact' nature of the model they consider, in their approach the CER on the short run dynamics of the system do not have the highly non-linear nature featured by NK-DSGE models which are prominent examples of 'inexact' rational expectation systems. Our approach extends the original idea of Johansen and Swensen (1999) to 'inexact' models and, as we show at the end of this section, to the case in which the linear rational expectations model features unobserved (latent) components.

The method discussed so far is based on the maintained assumption that the econometrician observes all components of Z_t and Y_t. We now turn to the role of unobservable components. An obvious way to face this issue and apply the 'LR1→LR2 →LR3' testing strategy is to appeal to 'additional' structural information related to the unobserved components in Y_t.[5] For instance, Bekaert et al. (2010) show how the information provided by the term structure of interest rates can be constructively incorporated and used to transform a standard NK-DSGE model featuring unobservable potential output and inflation target into a tractable linear system that can be estimated by 'standard' methods. Galí et al. (2001) follow a similar approach by assuming that under certain restrictions on technology and labour market structure, real marginal costs are proportionately related to the output gap within a local neighborhood of the steady state. We use a similar route in the empirical illustration of Section 6, where we approximate the output gap with the official measure provided by the Congressional Budget Office (CBO).

When proxies for the unobservables are not directly available and Assumption 2 is taken into explicit account, the 'LR1→LR2 →LR3' testing strategy can be adapted. As is known, and as the previous example show, while cointegration is invariant to extensions of the information set, it is not invariant to its reductions. This means that one can recover part of the cointegration implications of the model from observing Z_t^o (Y_t^o), as shown in the previous sub-sections. Yet, some of the long-run implications of the NK-DSGE model are still testable (see also Juselius, 2011).

In this case, the procedure is based on the following testing steps:

LR1 [Cointegration rank test: the case of unobservables] We specify a VAR model (with possibly with many lags) for the vector $Z_t^o = W_t^o$ (see Section 3) and test for the cointegration rank using the 'one shot' version of the LR Trace test.[6] If the cointegration rank implied by the NK-DSGE model—see Eq. (26) for the example—is rejected, we reject H_0 in Eq. (31) in favour of H_1 in Eq. (32). If instead the selected cointegration rank is model consistent we consider the next step.

LR2 [Overidentification cointegration restrictions test: the case of unobservables] We

---

[5]We refer to Fukač and Pagan (2010) for a discussion of the impact of off-model filtering upon the estimation of DSGE models.

[6]Our point is that the finite order VAR for Z_t^o with many lags should provide a reasonable approximation of the actual time series representation of Z_t^o, which is of VARMA-type under the null of valid NK-DSGE model, see Appendix B. Admittedly, a reasonable concern here is whether the cointegration rank test performed on the finte order VAR for of Z_t^o retains its usual properties. In principle, under certain conditions and with qualifications, the cointegration properties of the data should be invariant to the specification of the transient dynamics. Moreover, it is possible to apply a number of alternative cointegration rank tests, reviewed in e.g. Lütkepohl and Claessen (1997), which do not require estimating a fully identified VARMA-type model, see also Stock and Watson (1988).

compute the LR2 test by considering the cointegration implications of the NK-DSGE model subsumed in the vector $Z_t^o = W_t^o$. If the LR test rejects the over-identifying restrictions, we reject H_0 in Eq. (31) in favour of H_1 in Eq. (32), otherwise we consider the next step.

LR3 [**Test for CER: the case of unobservables**] We compute the LR3 test by evaluating the constrained likelihood of the transformed VAR model for Y_t by using a Kalman filter approach, see e.g. Ruge-Murcia (2007), Dave and DeJong (2007) and Fukač and Pagan (2010). If the CER are not rejected, H_0 is accepted and $\hat{\theta}^Y$ is the ML estimator of the structural parameters of the NK-DSGE model. In alternative to the LR3 test one can use, borrowing from the indirect inference literature, the cointegrated VAR for $Z_t^o = W_t^o$ not rejected in the previous two steps as the auxiliary model associated with the NK-DSGE system; this amounts to establishing a mapping from the NK-DSGE model parameters to the VAR parameters and then testing the data adequacy of the implied 'binding function' using the methods discussed in e.g. Smith (1993), Gourieroux et al. (1993) and Ruge-Murcia (2007).

Under Assumptions 1-2 and the null of correct specification H_0 in Eq. (31), the asymptotic properties of each of the three tests comprising the 'LR1→LR2 →LR3' testing strategy are known. The asymptotic properties of LR1 and LR2 may be found in Johansen (1996), while the asymptotic properties of LR3 are standard. Since under H_0 the three tests are correctly sized (and in particular their asymptotic size is equal to the nominal type-I error pre-fixed by the researcher), a simple Bonferroni argument suggests that the asymptotic size of the overall procedure does not exceed the sum of the type I error pre-fixed for the individual tests. This means that if the test for H_0 in Eq. (31) against H_1 in Eq. (32) is conducted by fixing the overall significance level at e.g. the 5% level, the critical values of the tests LR1, LR2 and LR3 must be chosen accordingly. Thus, if the practitioner wishes to test the null H_0 in Eq. (31) against the alternative H_1 in Eq. (32) at the 5% overall nominal level of significance, a reasonable choice is to fix the nominal significance level of the test LR1 at the 1% level, and the type-I error of the tests LR2 and LR3 at the 2% level.

We have postponed to Appendix A a more detailed derivation based on a refinement of the Bonferroni argument of the asymptotic size properties of the 'LR1→LR2 →LR3' testing strategy. The procedure will be consistent against all hypotheses with respect to which the individual tests LR1, LR2 and LR3 are consistent, including (i) DGPs in which the actual number of common stochastic trends is not the one implied by the NK-DSGE model, (ii) DGPs in which the number of common stochastic trends is the one predicted by the NK-DSGE model but the identification structure of the cointegration matrix is at odds with the requirements of the theoretical model, and (iii) DGPs in which the CER do not hold.

5 Simulation experiment

To evaluate the finite sample performance of the test sequence 'LR1→LR2 →LR3' under the null H_0 in Eq. (31), we conduct a small Monte Carlo experiment.

The set-up of the experiments is as follows. We assume that the DGP belongs to the family of determinate solutions associated with the NK-DSGE model for $Z_t = (y_t, \pi_t, i_t, y_t^p)$ discussed in the example of Section 3, reproduced here for completeness:

$$(y_t - y_t^p) = \gamma E \left(y_{t+1} - y_{t+1}^p \right) - \delta(i_t - E_t \pi_{t+1}) + (1 - \gamma) \left(y_{t-1} - y_{t-1}^p \right) + \eta_{\tilde{y},t} \tag{33}$$

$$\pi_t = \omega_f E_t \pi_{t+1} + \omega_b \pi_{t-1} + \kappa \left(y_t - y_t^p \right) + \eta_{\pi,t} \tag{34}$$

$$i_t = \rho i_{t-1} + (1 - \rho) \left[\varphi_\pi \pi_t + \varphi_y \left(y_t - y_t^p \right) \right] + \eta_{i,t} \tag{35}$$

$$y_t^p = y_{t-1}^p + \eta_{y^p,t} \tag{36}$$

$$\underbrace{\begin{pmatrix} \eta_{\tilde{y},t} \\ \eta_{\pi,t} \\ \eta_{i,t} \\ \eta_{y^p,t} \end{pmatrix}}_{\eta_t^Z} = \underbrace{\begin{pmatrix} \rho_{\tilde{y}} & 0 & 0 & 0 \\ 0 & \rho_\pi & 0 & 0 \\ 0 & 0 & \rho_i & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{R_Z} \underbrace{\begin{pmatrix} \eta_{\tilde{y},t-1} \\ \eta_{\pi,t-1} \\ \eta_{i,t-1} \\ \eta_{y^p,t-1} \end{pmatrix}}_{\eta_{t-1}^Z} + \underbrace{\begin{pmatrix} u_{\tilde{y},t-1} \\ u_{\pi,t-1} \\ u_{i,t-1} \\ u_{y^p,t-1} \end{pmatrix}}_{u_t^Z} \tag{37}$$

$$u_t^Z \sim WNN \left(0_{4\times1}, \Sigma_{Z,u} \right) \quad , \quad \Sigma_{Z,u} = \begin{pmatrix} \sigma_{\tilde{y}}^2 & 0 & 0 & 0 \\ 0 & \sigma_\pi^2 & 0 & 0 \\ 0 & 0 & \sigma_i^2 & 0 \\ 0 & 0 & 0 & \sigma_{y^p}^2 \end{pmatrix}. \tag{38}$$

The parameters ω_f and ω_b in the NKPC in Eq. (34) are restricted such that $\omega_f = \varrho/(1 + \varrho\varkappa)$, $\omega_b = \varkappa/(1 + \varrho\varkappa)$, where ϱ is the firms' discount factor which is kept fixed at the known value $\varrho = 0.99$, and $\varkappa$ captures the extent of firms' indexation to past prices, see Sub-section 2.1. The 'free' structural parameters are $\theta = (\gamma, \delta, \varkappa, \rho, \varphi_\pi, \varphi_y, \rho_{\tilde{y}}, \rho_\pi, \rho_i, \sigma_{\tilde{y}}^2, \sigma_\pi^2, \sigma_i^2)'$ and $\theta^e = (\theta', \theta^a)'$, where $\theta^a = Var(u_{y^p,t}) = \sigma_{y^p}^2$. The vector of fundamental shocks u_t^Z is assumed White Noise Gaussian with covariance matrix $\Sigma_{Z,u}$ above. The parameter vector θ is calibrated to the empirical estimates of Benati and Surico (2009), see in particular the last column of their Table 1 ('After the Volcker stabilization'), while the variance of potential output $\sigma_{y^p}^2$ is fixed at a value considered reasonable. The complete vector θ^e (θ^Y) chosen for this Monte Carlo experiment is reported in the left-most column of Panel 2 of our Table 1.

As shown in Section 4, the system (33)-(38) involves $I(1)$ variables and can be mapped into the representation in Eq.s (20)-(21) which is based on the stationary variables in Y_t. Thus, for a given θ^e (θ^Y) i.e. for given A_0^Y, A_f^Y, A_b^Y and $\Sigma_{Y,u}$ and fixed the initial conditions Y_0 and Y_{-1}, we generate the sequence $Y_1, ..., Y_T$ from the VAR system (27)-(30). We next use the

restriction $\beta_0 = \beta_0^b = \zeta'$ from Eq. (18) and the mapping $Z_t = G\left(\beta_0^b, \tau, 1 - L\right)^{-1} Y_t$ to obtain the observations $Z_1, ..., Z_T$. We then have all the ingredients to replicate M times the 'LR1→LR2 →LR3' procedure and evaluate the empirical sizes of the tests LR1, LR2 and LR3 individually and the overall empirical size, along with and the empirical performance of the ML estimator of θ^e.

[Table 1 about here.]

We generate time series of size $T = 100$, 200 and 500 from the NK-DSGE model $M = 10000$ times, and then compute the test sequence 'LR1→LR2 →LR3'.[7] For each replication, a sample of $T+200$ observations of Y_t (and then Z_t) is generated, and the first 200 observations are then discarded. To investigate the performance of our procedure, we use the overall nominal significance level of 5% (ψ =0.05), and consistent with this choice, we fix the nominal type-I errors of three tests of the procedure as follows: 1% for the LR1 test ($\psi_1 = 0.01$), 2% for the test LR2 ($\psi_2 = 0.02$) and 2% for the test LR3 ($\psi_3 = 0.02$). The corresponding critical values are chosen accordingly. The 1% critical values of the 'one-shot' cointegration rank test LR1 are the asymptotic ones derived by Doornik (1998), see also Juselius (2006, p. 419).

The likelihood maximization of the VAR system (27) under the constraints in Eq.s (28)-(30) is carried out using a numerical approximation of the non-linear CER. To simplify the computation burden, the likelihood maximization is carried out by treating ω_f as 'free' parameter, hence we replace $\varkappa$ with ω_f in the vector θ. Given the ML estimate of $\hat{\omega}_f$, we use the relationship $\omega_f = 0.99/(1 + 0.99\varkappa)$ to derive the indirect ML estimator of $\varkappa$, $\hat{\varkappa} = (1/\hat{\omega}_f - 1/0.99)$, and the restriction $\omega_b = (1/0.99)(1 - \omega_f/0.99)$ to impose the condition $\omega_f + \omega_b < 1$.

The results of this Monte Carlo experiment are summarized in Table 1. The empirical sizes of the individual tests and of the overall testing procedure are reported in Panel 1, while the Monte Carlo means and standard errors of the ML estimates of θ^e are reported in Panel 2. The rejection frequencies of the test statistics in Panel 1 are computed conditional on the non-rejection of the NK-DSGE model by the tests which come earlier in the sequence 'LR1→LR2 →LR3'. These conditional rejection frequencies sum up to the overall empirical size of the 'LR1→LR2 →LR3' procedure.

Before discussing the empirical size of the overall testing strategy, we discuss the empirical size of its components. The test LR1 is the 'one-shot' version of Johansen's LR trace test for cointegration rank and is conducted with H_r: $r = 3$ (one common stochastic trend), as predicted by the NK-DSGE model, against the alternative H_A: $r = n = 4$ (stationary system),

---

[7]All computations in this Monte Carlo experiment and in Section 6 where the NK-DSGE model is estimated using U.S. data have been computed in Ox.

using the 1% asymptotic critical value. The empirical size of the test is reported in the row labeled '$LR1\,(\psi_1 = 0.01)$: test of rank' in Panel 1. The empirical size of the LR1 test is not much affected by the number of observations in the sample and is between 0.6% and 1.1% for all sample sizes. A closer look at Table 1 reveals that LR1 test is under-sized for small samples with $T = 100$. This result is unexpected, as one would typically expect a much higher empirical size in samples of length $T = 100$ compared to the case $T = 500$, since many simulation studies have shown that the asymptotic critical values might be of little use in small samples. However, it is well known that the finite sample performance of the LR cointegration rank test may well depend on the structure of the short-run dynamics of the system which, in our set-up, is 'special', in the sense of being highly restricted by the CER.[8]

As concerns the test LR2, we recall that this is Johansen's 1991 likelihood ratio test of overidentified cointegrating vectors for a given rank. In our experiment, the identified cointegration matrix β_0 is fixed at the structure of the theoretical model in Eq. (22), $\beta_0 = \beta_0^b = \zeta$. The test is asymptotically distributed as chi-square with 3 degrees of freedom under the null. If the LR2 test rejects the over-identifying restrictions, the NK-DSGE model is rejected for the realizations of the DGP in the relevant replication. The empirical size of the LR2 test is reported in the second row of Panel 1 in Table 1, labeled '$LR2\,(\psi_2 = 0.02)$: test of beta'. As recognized in Johansen (2000, 2002) the limit distribution of the test is often a poor approximation to the finite sample distribution, and earlier simulation studies have shown that the empirical sizes tend to be much higher than nominal sizes, see *inter alia* Bewley et al. (1994), Fachin (2000), Gonzalo (1994), Li and Maddala (1997), Omtzigt and Fachin (2006). Our results add to this evidence, but the size distortions are not as bad as might be expected. For a sample length of $T = 100$, the empirical size is 6.7% as opposed to the 2% nominal size. However, the empirical size tends to be uniformly diminishing towards the nominal size of 2% as the sample size increases. For a sample size of $T = 500$, the rejection frequency of the LR2 test is 2.8%.

Finally, the test LR3 for the CER is reported in the third row of Panel 1 in Table 1, labeled '$LR3\,(\psi_3 = 0.02)$: test of CER'. The test is conducted conditional on the non-rejection of the beta matrix by LR2. The test statistic is compared with the critical value taken from the chi-square distribution with 28 degrees of freedom, the difference between the number of unrestricted parameters in the VAR (32+10) and the structural parameters ($\dim(\theta^e) = 14$). The empirical size is very good for samples of $T = 200$ and higher.

The empirical rejection frequency associated with the 'LR1→LR2 →LR3' testing strategy

---

[8]We notice for completeness that if for $T = 100$ one uses the asymptotic critical values taken from Table 15.1 in Johansen (1996) for the LR1 cointegration rank test in place of the critical values from Doornik (1998), the empirical size of the LR1 test turns out to be 0.015 (1.5%). We do not find significant differences in the results in samples of length $T = 200$ and $T = 500$.

is summarized in the seventh row of Panel 1 in Table 1 and can be compared to the overall empirical rejection frequency of the DSGE model reported in the eighth row. The performance of the 'LR1→LR2 →LR3' testing strategy obviously reflects the empirical size behavior already discussed of the three tests. It can be noticed that it ranges from 9.8% when $T = 100$ to 5.7% when $T = 500$, as opposed to a nominal type-I error of 5%. We deduce, therefore, that in samples of lengths typically available to practitioners, the use of small sample correction methods may improve the empirical performance of the testing strategy. For instance, albeit in this experiment the empirical performance of the cointegration rank test is satisfactory, in more general contexts it can be refined by using the methods recently proposed by Swensen (2006) and Cavaliere et al. (2012); similarly, the empirical performance of the LR2 test can be improved by using standard Bartlett corrections or bootstrap techniques, as advocated by Johansen (2000, 2002), Fachin (2000), Li and Maddala (1997), Omtzigt and Fachin (2006). As regards the test LR3, Cho and Moreno (2006) and Fanelli and Palomba (2011) have shown that bootstrap methods deliver reasonable size coverage in the presence of the highly non-linear CER implied by the rational expectations hypothesis.[9]

Panel 2 of Table 1 reports the Monte Carlo means of the structural parameters with the Monte Carlo standard errors in parentheses, i.e. the average of the i-th component of θ^e, θ_i^e, computed as $\hat{E}_{MC}\left(\hat{\theta}_i^e\right) = \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}_{ij}^e$, where $\hat{\theta}_{ij}^Y$ is ML estimate of θ_i^e obtained in the j-th simulation and $B < M$ is the number of DGPs for which the 'LR1→LR2 →LR3' test does not reject the NK-DSGE model; the associated Monte Carlo standard errors are calculated accordingly.

The structural parameters are recovered with surprising precision. The only exceptions, in samples of length $T = 100$, are the parameters of the policy rule φ_y and φ_π, although the estimation precision is increasing with the sample size. This lack of precision is a common finding and source of misunderstandings in the literature. In a recent influential paper, Cochrane (2011) has argued that the parameters of Taylor-type rules like that in Eq. (35) are not identifiable. Cochrane (2011), however, does not consider the 'hybrid' specification in Eq.s (33)-(37) but a less dynamic formulation of the NK-DSGE model. As it is known, identification problems in a system of variables featuring highly non-linear restrictions may involve the rank condition of the information matrix, or the relationship between the structural parameters and the sample objective function (in our case the likelihood function) which may display 'small' curvature in certain regions of the parameter space. The former concept of identification is also referred to as 'mathematical identification' (Johansen, 2010, p. 262) or 'population identification' (Canova

---

[9]We observe, however, that when the ML estimation is carried out through a grid search on all or part of the structural parameters (see Section 6), the implied computation burden may discourage the use of bootstrap techniques for the test LR3.

and Sala, 2009), as opposed to the latter, often termed 'sample identification', because it is specific to a particular data set and sample size. The results in Table 1 show that the chosen 'hybrid' NK-DSGE model is only moderately affected by 'sample identification' issues and this fact is confirmed by the empirical size of the LR3 test which for $T = 200$ is remarkably close to the the 2% nominal level.

In our Monte-Carlo experiment, the policy inertia parameter ρ is calibrated in the DGP to a relatively high value (0.834) and this fact might explain, in line with what argued by Mavroeidis (2010), why the coefficients $\varphi_{\tilde{y}}$ and φ_{π} of the reaction function are less accurately estimated in small samples.

[Figure 1 about here.]

[Figure 2 about here.]

That the difficulty of identifying the policy parameters φ_y and φ_{π} reflects a small sample issue rather than a 'population identification' issue is further confirmed by the graphs we have reported in Figures 1 and 2. Here we have plotted the marginal empirical distributions of the ML estimators of some of the elements in the vector θ^e obtained from the Monte Carlo experiment, for the cases $T = 100$ and $T = 500$, respectively. We notice that the sample distributions of $\hat{\varphi}_y$ and $\hat{\varphi}_{\pi}$ tend to be more concentrated around their 'true' values as the sample size increases. Instead, we observe that the marginal sample distributions of the ML estimator of the forward-looking parameter of the NKPC, ω_f, displays a substantial bimodality which does not disappear in samples of length $T = 500$ (recall that $\omega_f = 0.99/(1 + 0.99\varkappa)$ and that in our Monte Carlo experiment we estimated ω_f freely and $\varkappa$ indirectly). The graphs also confirm that difficulties that characterize the estimation of the slope parameter of the NKPC, κ, must be ascribed to small sample issues. We leave a detailed investigation of these interesting issues, which perhaps helps to explain some controversial results about the estimation of the NKPC in the literature, to future research.

For large samples, the test sequence 'LR1→LR2 →LR3' seems to work well. For small samples, the overall size is somewhat distorted by the well-known small sample problems of the LR2 test.

Keeping these results in mind, we next turn to an empirical application of our test procedure.

6   An estimated NK-DSGE model of the U.S. economy

When all variables W_t (not to be confounded with the case of Z_t) of our reference small NK-DSGE model are observed, the 'LR1→LR2 →LR3' testing strategy discussed in this paper can be

properly adapted as illustrated in this section. We estimate the reference NK-DSGE monetary model summarized in Eq.s (1)-(2) using U.S. quarterly data, approximating potential output with the official measure provided by the Congressional Budget Office (CBO) estimation like, *inter alia*, Cho and Moreno (2006) and Castelnuovo and Fanelli (2011). This solution allows us not to consider Assumption 2 because we treat the output gap as an observed variable. Thus, we take system (1)-(2) directly to the data.

Contrary to what we have done in the Monte Carlo experiment, we treat the indexation parameter $\varkappa$ of the NKPC as a 'free' parameters (for reasons that will be clear below) and estimate ω_f and ω_b accordingly. Moreover, differently from Benati and Surico (2009), we do not force the covariance matrix of structural disturbances, $\Sigma_{W,u}$, to be diagonal, see e.g. Dufour et al. (2009) and Castelnuovo and Fanelli (2011) for similar choices. We split the vector θ as $\theta = (\theta'_s, \theta'_u)'$, where $\theta_s = (\gamma, \delta, \varrho, \varkappa, \rho, \varphi_\pi, \varphi_y, \rho_{\tilde{y}}, \rho_\pi, \rho_R)'$ and $\theta_u = vech(\Sigma_{W,u})$.

We employ quarterly data relative to the 'Great Moderation' sample 1985q1-2008q3. Four arguments motivate our choice: (i) the 'credibility build-up' undertaken by the Federal Reserve in the early 1980s, a period during which private agents gradually changed their view on the Fed's ability to deliver low inflation (Goodfriend and King, 2005); (ii) the first years of Volcker's tenure (until October 1982) were characterized by non-borrowed reserves targeting, hence one can hardly expect a good fit of conventional policy rules within this period, a fact that would carry consequences on the estimates of all parameters of the system, see Mavroeidis (2010) and references therein; (iii) the end of the sample 2008q3 is justified by our intention to avoid dealing with the 'zero-lower bound' phase began in December 2008, which triggered a series of non-standard policy moves by the Federal Reserve; (iv) formal testing analysis by Castelnuovo and Fanelli (2011) shows that the reference NK-DSGE model in Eq.s (10)-(13) has a unique stable solution over the 1985q1-2008q3 period, while the picture is more controversial if other sample periods are considered, and our approach requires the system to be in a determinate state.

The variables used in the empirical analysis are real GDP, GDP_t; the CBO measure of potential output, GDP_t^p; the inflation rate π_t which is the quarterly growth rate of the GDP deflator; the short-term nominal interest rate, i_t, given by the effective Federal funds rate expressed in quarterly terms (averages of monthly values). The output gap $\tilde{y}_t$ is computed as percent log-deviation of the real GDP with respect to the CBO potential output: $\tilde{y}_t = y_t - y_t^p = \log(GDP_t) - \log(GDP_t^p) \simeq \left(\frac{GDP_t}{GDP_t^p} - 1 \right)$. The source of the data is the Federal Reserve Bank of St. Louis' web site.

The assumption that GDP_t^p is proxied with an observable time series allows us to treat the

vector $W_t = (\tilde{y}_t, \pi_t, i_t)' = W_t^o$ as observable.[10] The unique asymptotically stable solution of the NK-DSGE model is given by the constrained VAR model in Eq. (3), where the reduced form coefficients $F_1 = F_1(\theta)$, $F_2 = F_2(\theta)$ and $Q = Q(\theta)$ depend non-linearly on the structural parameters θ through the CER in Eq.s (4)-(6). The ML estimation of θ is obtained by taking system (3) directly to the data.

In this set-up, the null hypothesis under investigation is

H_0': the DGP belongs to the VAR solution in Eq.s (3)-(6) of the NK-DSGE model

and the alternative is

H_1': the DGP is not consistent with the VAR solution in Eq.s (3)-(6).

The 'LR1→LR2 →LR3' testing strategy collapses to the 'LR1&2→LR3' sequence, whose steps are described in the following.

LR1&2 [stationary W_t^o system] Estimate the unrestricted counterpart of the VAR model for $W_t^o = (\tilde{y}_t, \pi_t, i_t)'$, see Eq. (3), and test for the stationarity of W_t^o. If stationarity is rejected, H_0' is rejected, otherwise consider the next test.

LR3 [Test for CER] Estimate θ from the VAR representation in Eq.s (3)-(6) by ML, obtaining $\hat{\theta} = (\hat{\theta}_s', \hat{\theta}_u')'$. Then compute a LR test for the CER: if the CER are rejected, H_0' is rejected, otherwise the NK-DSGE model is supported by the data.

Thus, in this simplified set-up, we have simply 'merged' the tests LR1 and LR2 in a single LR test which assesses the stationarity of the VAR for W_t^o.

[Table 2 about here.]

Our empirical analysis starts with the estimation of an unrestricted VAR system for $W_t^o = (\tilde{y}_t, \pi_t, i_t)'$ with two lags (henceforth VAR(2)) on the 1985q1-2008q3 period. Estimation is performed by including a constant in the equations because the variables in W_t^o were not demeaned prior to estimation. The upper panel of Table 2 reports the estimated unrestricted reduced form VAR coefficients, while the lower panel summarizes some diagnostic tests, including a test for the absence of autocorrelation in the disturbances, a test for the absence of ARCH-type

---

[10]Obviously, the econometric analysis could be also based on the system $Z_t = (y_t, \pi_t, i_t, GDP_t^p)'$, but the information set is less well suited to model potential output. Another alternative would be to treat potential output as exogenous and use the approach of Harbo et al. (1998). The results of these alternative approaches are available upon request.

components in the disturbances and a test for the hypothesis of Gaussian disturbances. Table 3 reports the eigenvalues of the companion matrix of the unrestrictedly estimated VAR(2) and the LR cointegration rank test for the hypothesis that there are two cointegrating relations in the system (i.e. one common stochastic trend) against the alternative of a stationary VAR. This LR test acts as the LR1&2 test described above.

[Table 3 about here.]

Overall, the results in Tables 2-3 suggest that a stationary VAR representation for the variables in $W_t^o = (\tilde{y}_t, \pi_t, i_t)'$ stands as a reasonably good approximation of U.S. quarterly data over the period 1985.q1-2008.q3. We will turn on the interpretation of the LR1&2 test reported in the bottom part of Table 3 at the end of this section.

We then proceed with the estimation of the structural parameters $\theta = (\theta_s', \theta_u')'$ by maximizing the log-likelihood of the VAR(2) under the CER in Eqs. (4)-(6). We use a grid search for the parameters δ, ϱ, $\varkappa$, φ_y and φ_π, which are notoriously difficult to estimate through non-Bayesian techniques. Estimation results are summarized in the upper panel of Table 4, while the test LR3 for the CER is reported in the lower panel.

[Table 4 about here.]

Our point estimates turn out to be quite similar to those in a variety of contributions in the literature, hence we do not discuss these results in details. Nevertheless, a note of caution is needed for the policy parameters: as suggested by the Monte Carlo section, with about 100 observations it is extremely difficult to obtain precise estimates of the policy reaction function $\varphi_{\tilde{y}}$ and φ_π; the same is true for the slope of the NKPC κ and the indexation parameter $\varkappa$. Our results confirm this evidence and the one recently reported in Mavroeidis (2010).

If the overall type-I error for the null hypothesis that the NK-DSGE model is valid (our H_0') is pre-fixed at the 5% level, a reasonable choice for our sequential testing procedure is the 2.5% level for the test LR1&2 and the 2.5% level for the test LR3. The results in Table 4 show that, quite surprisingly, that LR1&2 accepts the stationarity of the system while LR3 rejects the CER, albeit only marginally. Unfortunately, we can not easily provide a bootstrap version of the test LR3 because the grid search procedure used to maximize the likelihood function is computationally intensive and the computation burden would be relevant. Nonetheless, the indication emerging from our analysis is that the overall set of restrictions implied by the NK-DSGE model is rejected only marginally, hence the model is not completely at odds with the data on the period 1985q1-2008q3.[11]

---

[11]Our testing result has been obtained by specifying the covariance matrix of fundamental disturbances, $\Sigma_{W,u}$,

7 Lessons for practitioners

In this paper we have proposed a new approach to evaluate the empirical reliability of the class of NK-DSGE models used in monetary policy analysis based on the idea of testing jointly all restrictions these models places on their unique stable solution under rational expectations. We adopt a likelihood-based perspective and consider the set of restrictions at low and high frequencies which the NK-DSGE model places on its reduced form VAR solution. The novelty of our evaluation method is that the empirical assessment of the NK-DSGE model is based on a conditional sequence of likelihood-ratio tests which assess the long-run and short-run features of the model jointly and through which we can control at which stage of the process the system is rejected by the data. Our approach is logically comparable, with many qualifications discusses throughout the paper, with the methodology originally proposed by Campbell and Shiller (1987) for estimating and testing present value models through VAR systems. Furthermore, it complements the evaluation approach suggested by Fukač and Pagan (2010) in a 'limited-information' framework and generalizes the seminal method advocated by Canova et al. (1994) to evaluate real business cycle models to a wider set-up.

Our analysis, based on a simulation experiment and the estimation of a monetary NK-DSGE model using U.S. data, reveals that some conclusions about the frequentist approach to the estimation of NK-DSGE models can be reached.

First, the empirical evaluation of a NK-DSGE model should be carried out by considering all long-run and short-run restrictions, not just a sub-set of them.

Second, even though our procedure consists of several tests, the size of the overall procedure is under control. The 'LR1→LR2 →LR3' testing strategy has been explicitly designed to assess the empirical validity of the joint set of restrictions implied by the NK-DSGE model at a pre-fixed significance level. Our simulation experiment shows that despite the highly non-linear nature of the CER, the overall empirical size can be controlled.

Third, the size control of the proposed testing strategy can be improved by various methods. In particular, the test for steady-state restrictions, LR2, is probably best performed using a small sample adjustment, in the form of a Bartlett correction or a bootstrap procedure, see in particular the evidence presented in Omtzigt and Fachin (2006).

---

non-diagonal. Indeed, in Table 4 we do find significant off-diagonal terms in the estimated $\Sigma_{W,u}$ matrix. This result can be interpreted by observing that from Eq. (2) one can derive the relationship

$$vec(\Sigma_{W,u}) = [(I_p \otimes I_p) - (R_W \otimes R_W)] \, vec(\Sigma_{W,\eta})$$

where $\Sigma_{W,\eta}$ is the covariance matrix of the disturbances η_t^W of the NK-DSGE model: the significant cross-correlations found in $\hat{\Sigma}_{W,u}$ might simply reflect a non-diagonal structure of the R_W matrix in the DGP. Interestingly, with both R_W and $\Sigma_{W,u}$ diagonal, we would have rejeced the NK-DSGE model by the test LR3.

Fourth, ML estimation is involved for two reasons. On the one hand, it requires a numerical approximation of the non-linear CER; the structural parameters are subject to bound (other than sign) constraints whose omission gives rise to the so-called 'absurd values' phenomenon, on the other hand. These arguments, however, are not sufficient to discourage the use of ML estimation. The use of grid search for the parameters which are notoriously difficult to pick out from the data is a solution which, albeit the computational burden involving, guarantees that all restrictions are met, and allows one to retain, if certain conditions are met, some of the desirable properties of ML estimation.

Fifth, the claim that the parameters associated with the policy rule of a NK-DSGE model are unidentified (in the sense of not being associated with a singular information matrix) is false for 'hybrid' specifications. Obviously, in samples of lengths typically available to practitioners, weak identification (the difficulty of estimating these parameters precisely also under the null) is a concern that deserves attention; reliable inference in NK-DSGE models requires identification-robust methods as suggested by e.g. Guerron-Quintana et al. (2013), among many others.

Finally, our empirical application has shown that, despite the highly constrained nature of the model, a statistical evaluation of NK-DSGE models that account for the time series properties of the variables does not necessarily lead to rejection. The estimation conducted using U.S. quarterly data (and some simplyfying hypotheses) shows that the rejection of the monetary policy NK-DSGE model used in the recent literature is only marginal. In light of the recent developments in business cycle dynamics and the conduct of monetary policy, it will be crucial to re-evaluate the model as soon as enough data become available, especially if a paradigm that incorporates 'new' ingredients within the baseline framework will emerge.

References

An, S. and F. Schorfheide (2007). Bayesian analysis of DSGE models. *Econometric Reviews*, *26*(2-4), 113–172.

Bekaert, G., S. Cho and A. Moreno (2010). New Keynesian macroeconomics and the term structure. *Journal of Money Credit and Banking*, *42*, 33–61.

Benati, L. and P. Surico (2009). VAR analysis and the great moderation. *American Economic Review*, *99*(4), 1636–1652.

Bewley, R., D. Orden, M. Yang and L. A. Fisher (1994). Comparison of Box-Tiao and Johansen canonical estimators of cointegrating vectors in VEC(1) models. *Journal of Econometrics*, *64*, 3–27.

Binder, M. and M. H. Pesaran (1995). Multivariate rational expectations models and macroeconomic modelling: A review and some new results. In Pesaran, M. and M. Wickens (eds.), *Handbook of Applied Econometrics,*, 139–187. Blackwell Publishing Ltd, Oxford.

Campbell, J. and R. Shiller (1987). Cointegration and tests of present-value models. *Journal of Political Economy*, *95*, 1052–1088.

Canova, F. and F. Ferroni (2012). The dynamics of US inflation: Can monetary policy explain the changes? *Journal of Econometrics*, *167*(1), 47–60.

Canova, F., M. Finn and A. Pagan (1994). Evaluating a real business cycle model. In Hargreaves, C. P. (ed.), *Nonstationary Time Series Analysis and Cointegration*, Advanced Texts in Econometrics, 225–255. Oxford University Press, Oxford.

Canova, F. and L. Sala (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, *56*(4), 431 – 449.

Carlstrom, C. T., T. S. Fuerst and M. Paustian (2009). Monetary policy shocks, Choleski identification, and DNK models. *Journal of Monetary Economics*, *56*(7), 1014 – 1021.

Castelnuovo, E. and L. Fanelli (2011). Monetary policy indeterminacy in the U.S.: Results from a classical test. Working paper, Quaderni di Dipartimento, Serie Ricerche N. 8, University of Bologna.

Cavaliere, G., A. Rahbek and A. Taylor (2012). Boostrap determination of the co-integration rank in Vector Autoregressive models. *Econometrica*, *80*, 1721–1740.

Cho, S. and A. Moreno (2006). A small-sample study of the New-Keynesian macro model. *Journal of Money Credit and Banking*, *38*, 1462–1482.

Cochrane, J. H. (2011). Determinacy and identification with Taylor rules. *Journal of Political Economy*, *119*(3), pp. 565–615.

Cogley, T. (2001). Estimating and testing rational expectations models when the trend specification is uncertain. *Journal of Economic Dynamics and Control*, *25*(10), 1485–1525.

Dave, C. and D. DeJong (2007). *Structural Macroeconometrics*. Princeton University Press.

De Grauwe, P. (2010). The scientific foundation of dynamic stochastic general equilibrium (DSGE) models. *Public Choice*, *144*, 413–443.

Dees, S., M. H. Pesaran, H., L. V. Smith and R. P. Smith (2009). Identification of new Keynesian Phillips curves from a global perspective. *Journal of Money, Credit and Banking*, *41*(7), 1481–1502.

Del Negro, M., F. Schorfheide, F. Smets and R. Wouters (2007). On the fit of new Keynesian models. *Journal of Business & Economic Statistics*, *25*(2), 123–143.

Doornik, J. A. (1998). Approximations to the asymptotic distributions of cointegration tests. *Journal of Economic Surveys*, *12*(5), 573–593.

Dufour, J.-M., L. Khalaf and M. Kichian (2009). Structural multi-equation macroeconomic models: Identification-robust estimation and fit. Bank of Canada, Working Paper 2009-10.

Fachin, S. (2000). Bootstrap and asymptotic tests of long-run relationships in cointegrated systems. *Oxford Bulletin of Economics and Statistics*, *62*(4), 543–551.

Fanelli, L. (2008). Testing the new Keynesian Phillips curve through Vector Autoregressive models: Results from the Euro area. *Oxford Bulletin of Economics and Statistics*, *70*, 53–66.

Fanelli, L. (2011). Robust identification conditions form determinate and indeterminate Linear Rational Expectations models. Dipartimento di Scienze Statistiche, Quaderno di Dipartimento Serie Ricerche 1, University of Bologna.

Fanelli, L. (2012). Determinacy, indeterminacy and dynamic misspecification in linear rational expectations models. *Journal of Econometrics*, *170*(1), 153–163.

Fanelli, L. and G. Palomba (2011). Simulation-based tests of forward-looking models under VAR learning dynamics. *Journal of Applied Econometrics*, *26*, 762–782.

Fernandez-Villaverde, J., J. Rubio-Ramirez, T. J. Sargent and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review*, *97*, 1021–1026.

Franchi, M. and P. Paruolo (2012). DSGE models and cyclical co-movements in VARs. Working Paper.

Fukač, M. and A. Pagan (2010). Limited information estimation and evaluation of DSGE models. *Journal of Applied Econometrics*, *25*(1), 55–70.

Galí, J., M. Gertler and J. Lopez-Salido (2001). European inflation dynamics. *European Economic Review*, *45*, 1237–1270.

Gonzalo, J. (1994). Five alternative methods of estimating long run equilibrium relationships. *Journal of Econometrics*, *60*, 1–31.

Goodfriend, M. and R. G. King (2005). The incredible Volcker disinflation. *Journal of Monetary Economics*, *52*(5), 981–1015.

Gorodnichenko, Y. and S. Ng (2010). Estimation of DSGE models when the data are persistent. *Journal of Monetary Economics*, *57*(3), 325–340.

Gourieroux, C., A. Monfort and E. Renault (1993). Indirect inference. *Journal of Appled Econometrics*, *8*, 85–118.

Guerron-Quintana, P., A. Inoue and L. Kilian (2013). Frequentist inference in weakly identified DSGE models. *Quantitative Economics*, *forthcoming*.

Hannan, E. and M. Deistler (1988). *The statistical theory of linear systems*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley.

Hansen, L. P. and T. J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control*, *2*, 7–46.

Hansen, L. P. and T. J. Sargent (1981). Linear rational expectations models for dynamically interrelated variables. In Lucas, R. and T. Sargent (eds.), *Rational Expectations and Econometric Practise*, 127–156. University of Minnesota Press, Minneapolis.

Hansen, L. P. and T. J. Sargent (1991). Exact linear rational expectations models: Specification and estimation. In *Rational Expectations Econometrics*. Westview Press, Boulder.

Harbo, I., S. Johansen, B. Nielsen and A. Rahbek (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics*, *16*, 388–399.

Hendry, D. F. and G. E. Mizon (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, 272–300. Basil Blackwell, Oxford.

Ireland, P. (2004). Technology shocks in the new Keynesian model. *Review of Economics and Statistics*, *86*, 923–936.

Iskrev, N. (2008). Evaluating the information matrix in linearized DSGE models. *Economic Letters*, *99*, 607–610.

Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics*, *57*, 189–202.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, *59*, 1551–1580.

Johansen, S. (1996). *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model*. Oxford University Press, Oxford, 2nd edn.

Johansen, S. (2000). A Bartlett correction factor for tests on the cointegrating relations. *Econometric Theory*, *16*(5), 740–778.

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica*, *70*(5), 1929–1961.

Johansen, S. (2006). Confronting the economic model with the data. In Colander, D. (ed.), *Post Walrasian Macroeconomics*. Cambridge University Press.

Johansen, S. (2010). Some identification problems in the cointegrated vector autoregressive model. *Journal of Econometrics*, *158*, 262–273.

Johansen, S. and A. R. Swensen (1999). Testing exact rational expectations in cointegrated vector autoregressive models. *Journal of Econometrics*, *93*, 73–91.

Juselius, K. (2006). *The Cointegrated VAR model: Theory and Applications*. Oxford University Press, Oxford.

Juselius, K. and M. Franchi (2007). Taking a DSGE model to the data meaningfully. *Economics: The Open-Access, Open-Assessment E-Journal*, *1*(2007-4).

Juselius, M. (2011). Testing steady-state restrictions of linear rational expectations models when data are highly persistent. *Oxford Bulletin of Economics and Statistics*, *73*(3), 315–334.

King, R. G., C. I. Plosser, J. H. Stock and M. W. Watson (1991). Stochastic trends and economic fluctuations. *The American Economic Review*, *81*(4), 819–840.

Li, H. (2007). Small-sample inference in rational expectations models with persistent data. *Economic Letters*, *95*, 203–210.

Li, H. and G. S. Maddala (1997). Bootstrapping cointegrating regressions. *Journal of Econometrics*, *80*, 297–318.

Lubik, A. and F. Schorfheide (2004). Testing for indeterminacy: An application to US monetary policy. *American Economic Review*, *94*, 190–217.

Lubik, T. A. and F. Schorfheide (2003). Computing sunspot equilibria in linear rational expectations models. *Journal of Economic Dynamics and Control*, *28*(2), 273–285.

Lütkepohl, H. and H. Claessen (1997). Analysis of cointegrated VARMA processes. *Journal of Econometrics*, *80*, 223–239.

Mavroeidis, S. (2010). Monetary policy rules and macroeconomic stability: Some new evidence. *American Economic Review*, *100*, 491–503.

Omtzigt, P. and S. Fachin (2006). The size and power of bootstrap and Bartlett-corrected tests of hypotheses on the cointegrating vectors. *Econometric Reviews*, *25*(1), 41–60.

Pesaran, M. H. and R. Smith (2011). Beyond the DSGE straitjacket. *The Manchester School*, *Supplement*, 1–38.

Ravenna, F. (2007). Vector autoregressions and reduced form representations of DSGE models. *Journal of Monetary Economics*, *54*, 2048–2064.

Ruge-Murcia, F. (2007). Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, *31*, 2599–2636.

Schorfheide, F. (2000). Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics*, *15*, 645–670.

Schorfheide, F. (2011). Estimation and evaluation of DSGE models: progress and challenges. *NBER, Working Paper 16781*.

Smith, A. A., Jr. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, *8*, S63–S84.

Spanos, A. (2011). Foundational issues in statistical modeling: Statistical model specification and validation. *Rationality, Markets and Morals*, *2*, 146–178.

Stock, J. and M. Watson (1988). Testing for common trends. *Journal of the American Statistical Association*, *83*, 10971107.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica*, *74*, 1699–1714.

Uhlig, H. (1999). A toolkit for analyzing nonlinear dynamic stochastic models easily. In *Computational Methods for the Study of Dynamic Economies*. Ramon Marimon and Andrew Scott (editors).

Vredin, A. and P. Söderlind (1996). Applied cointegration analysis in the mirror of macroeconomic theory. *Journal of Applied Econometrics*, *11*, 363–382.

Wallis, K. (1980). Econometric implications of the rational expectations hypothesis. *Econometrica*, *48*, 49–73.

Appendix A   Size properties of the testing strategy

To fully understand the mechanics of testing evaluation procedure of the NK-DSGE model, in this Appendix we derive the asymptotic size properties of the 'LR1→LR2 →LR3' sequence.[12]

Denote with $LR_{i,T}$, $i = 1, 2, 3$, the three LR test statistics and let ψ_i be the nominal significance level (type-I error) pre-fixed for the i-th test; moreover, let $\psi_{i,T} = P_{i,T}^{H_0}(LR_{i,T} \geq cr_{i,T}^{\psi_i})$ be the exact size of the i-th test based on a sample of length T, where $P_{i,T}^{H_0}(\cdot)$ is the probability measure associated with the null distribution of $LR_{i,T}$ and $cr_{i,T}^{\psi_i}$ is the corresponding critical value at nominal level ψ_i. Under H_0 the three tests, individually considered, are correctly sized and in particular such that

$$\psi_{i,\infty} = \limsup_{T \to \infty} \psi_{i,T} = \psi_i \quad , \quad i = 1, 2, 3 \tag{A39}$$

where $\psi_{i,\infty}$ is the asymptotic size of the i-th test. Let $P_{1,2,T}^{H_0}(\cdot \; ; \; \cdot)$ and $P_{2,3,T}^{H_0}(\cdot \; ; \; \cdot)$ be the probability measures associated with the joint null distributions of the test statistics $LR_{1,T}$ and $LR_{2,T}$ and the test statistics $LR_{2,T}$ and $LR_{3,T}$, respectively. It turns out that the overall asymptotic size of the test for H_0 in Eq. (31) is given by

$$\psi_{\infty} = \limsup_{T \to \infty} \psi_T \tag{A40}$$

where

$$\psi_T = P_{1,T}^{H_0}(LR_{1,T} \geq cr_{1,T}^{\psi_1}) + P_{1,2,T}^{H_0}(LR_{1,T} < cr_{1,T}^{\psi_1} \; ; \; LR_{2,T} \geq cr_{2,T}^{\psi_2})$$
$$+ P_{2,3,T}^{H_0}(LR_{2,T} < cr_{2,T}^{\psi_2} \; ; \; LR_{3,T} \geq cr_{3,T}^{\psi_3}). \tag{A41}$$

The first addend of Eq. (A41) captures the probability, on a sample of length T, that the LR1 test incorrectly rejects the cointegration rank; the second addend captures the joint probability that the LR2 test incorrectly rejects the structure of the cointegration matrix and the LR1 test correctly selects the cointegration rank and, finally, the last addend captures the joint probability that the LR3 test incorrectly rejects the CER and the LR2 correctly rejects the structure of the cointegration matrix.

By using the inequalities

$$P_{1,2,T}^{H_0}(LR_{1,T} < cr_{1,T}^{\psi 1}; LR_{2,T} \geq cr_{2,T}^{\psi 2}) \leq \psi_{2,T}$$

$$P_{1,2,3,T}^{H_0}(LR_{2,T} < cr_{2,T}^{\psi_2} \; ; \; LR_{3,T} \geq cr_{3,T}^{\psi_3}) \leq \psi_{3,T},$$

---

[12]We refer to Spanos (2011) for a comprehensive treatment of size in general-to-specific sequential testing procedures.

the limit in Eq.s ([A40])-([A41]) is such that

$$\psi_\infty \leq \psi_{1,\infty} + \psi_{2,\infty} + \psi_{3,\infty} = \psi_1 + \psi_2 + \psi_3 \tag{A42}$$

hence the asymptotic size of the overall test sequence does not exceed the sum of the type I error pre-fixed for the individual tests. This result suggests that in empirical analyses it is convenient to fix the overall nominal significance level of the procedure at $\psi = (\psi_1 + \psi_2 + \psi_3)$.

Appendix B A, B, C (and D's) representation of the reference NK-DSGE model

Given the representation in Eq. (3) of our NK-DSGE model, we start from the state-space (companion) representation

$$\underbrace{\left(\begin{array}{c} W_t \\ W_{t-1} \end{array} \right)}_{x_t} = \underbrace{\left(\begin{array}{cc} \tilde{F}_1 & \tilde{F}_2 \\ I_n & 0_{n\times n} \end{array} \right)}_{A} \underbrace{\left(\begin{array}{c} W_{t-1} \\ W_{t-2} \end{array} \right)}_{x_{t-1}} + \underbrace{\left(\begin{array}{c} \tilde{Q} \\ 0_{n\times n} \end{array} \right)}_{B} u_t^W \qquad (B43)$$

where the matrices $\tilde{F}_1 = \tilde{F}_1(\theta)$, $\tilde{F}_2 = \tilde{F}_2(\theta)$ and $\tilde{Q} = \tilde{Q}(\theta)$ depend non-linearly on the structural parameters θ through the CER in Eqs. (4)-(6). A convenient way to summarize system (B43) is

$$\underset{2p\times 1}{x_t} = \underset{2p\times 2p}{A} \underset{2p\times 1}{x_{t-1}} + \underset{2p\times p}{B} \underset{p\times 1}{\varepsilon_t} \qquad (B44)$$

where the definition of the vectors x_t and $\varepsilon_t = u_t^W$ and of the matrix $A = A(\theta)$ is obvious. We notice that in this case the covariance matrix of the error term $v_t = B\varepsilon_t$ is singular. Let y_t be the $m \times 1$ vector that contains the endogenous observable variables; a natural choice for the measurement equation is given by

$$\underset{m\times 1}{y_t} = \underset{m\times 2p}{H} x_t \qquad (B45)$$

where H is a (known) $m \times 2p$ selection matrix that picks out the endogenous observable variables from x_t, see e.g. Dave and DeJong (2007), Iskrev (2008) and Iskrev (2010).

The system given by Eq.s (B44)-(B45) represents a conventional state-space representation of the NK-DSGE model and under the assumption that $\varepsilon_t = u_t^W$ is Gaussian, the Kalman filter can be used to build and evaluate the likelihood function, see, among many others, Ruge-Murcia (2007) and Dave and DeJong (2007). However, depending on the model at hand, solution method one uses to arrive from the structural equations to the representation in Eq. (B44), other choices for x_t and y_t are equally possible, see e.g. Uhlig (1999).

By using Eq. (B44) in Eq. (B45) yields

$$y_t = HAx_{t-1} + HB\varepsilon_t$$

which for $C = HA$ and $D = HB$ reads as

$$y_t = Cx_{t-1} + D\varepsilon_t. \qquad (B46)$$

The system obtained by coupling Eq. (B44) and Eq. (B46), reported here for convenience

$$x_t = Ax_{t-1} + B\varepsilon_t$$

$$y_t = Cx_{t-1} + D\varepsilon_t$$

captures the determinate equilibrium of the NK-DSGE model and defines the so-called 'A, B, C (and D's)' representation, see Fernandez-Villaverde et al. (2007), Ravenna (2007) and Franchi and Paruolo (2012); see also Hannan and Deistler (1988). For the case $m = p$ (D square) and D non-singular, Fernandez-Villaverde et al. (2007) and Ravenna (2007) discuss conditions under which y_t has a reduced form VAR (VAR(∞)) representation with ε_t as input variable; Franchi and Paruolo (2012) extend the analysis to the case of D non square and possibly singular. Given our NK-DSGE model and the definition of y_t in Eq. (B46), we face the case $m < p$, hence the results in Fernandez-Villaverde et al. (2007) and Ravenna (2007) can not be applied. A reasonable conjecture is that y_t has a finite order VARMA-type reduced form representation, as shown in e.g. Bekaert et al. (2010).
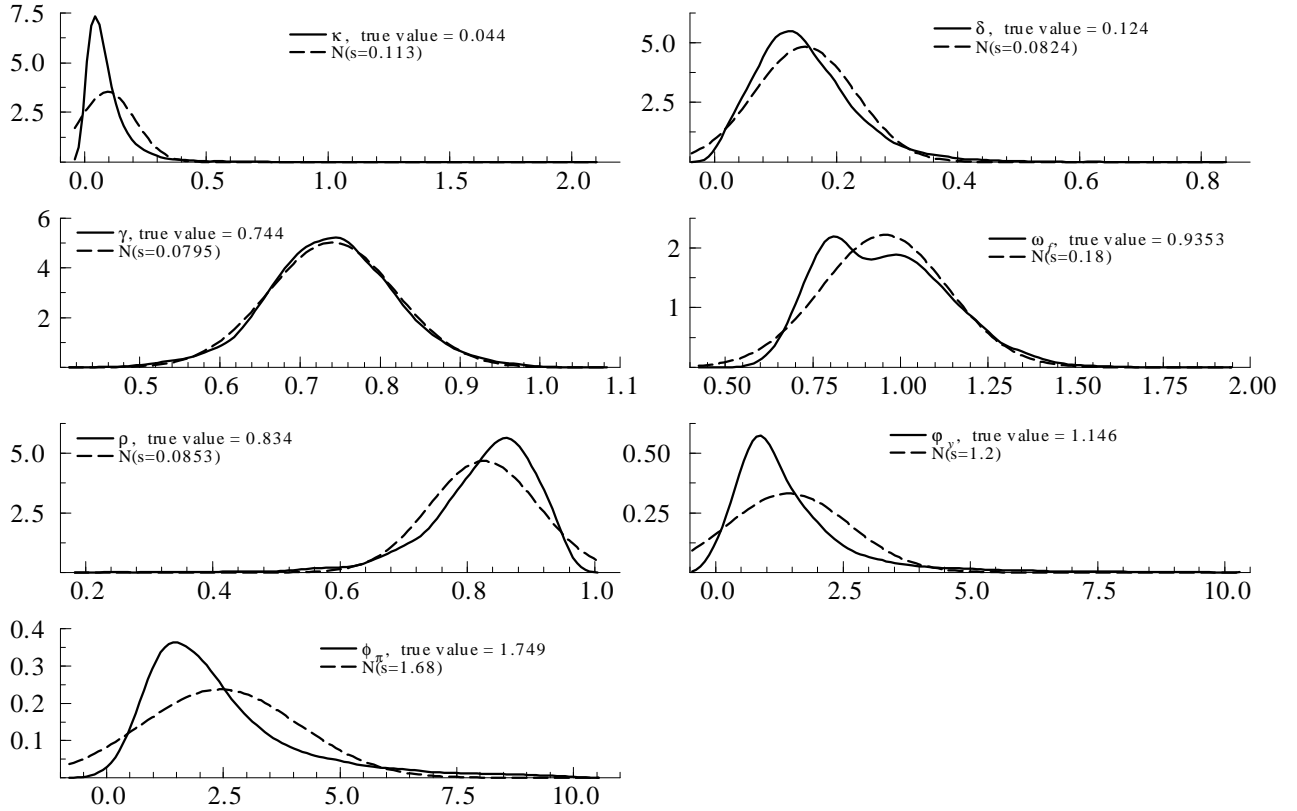
List of Figures

Figure 1: Empirical marginal distribution of the ML estimators of some parameters of the NK-DSGE model (solid lines) obtained from the Monte Carlo experiment against the Gaussian density (dotted lines) in samples of length $T = 100$.
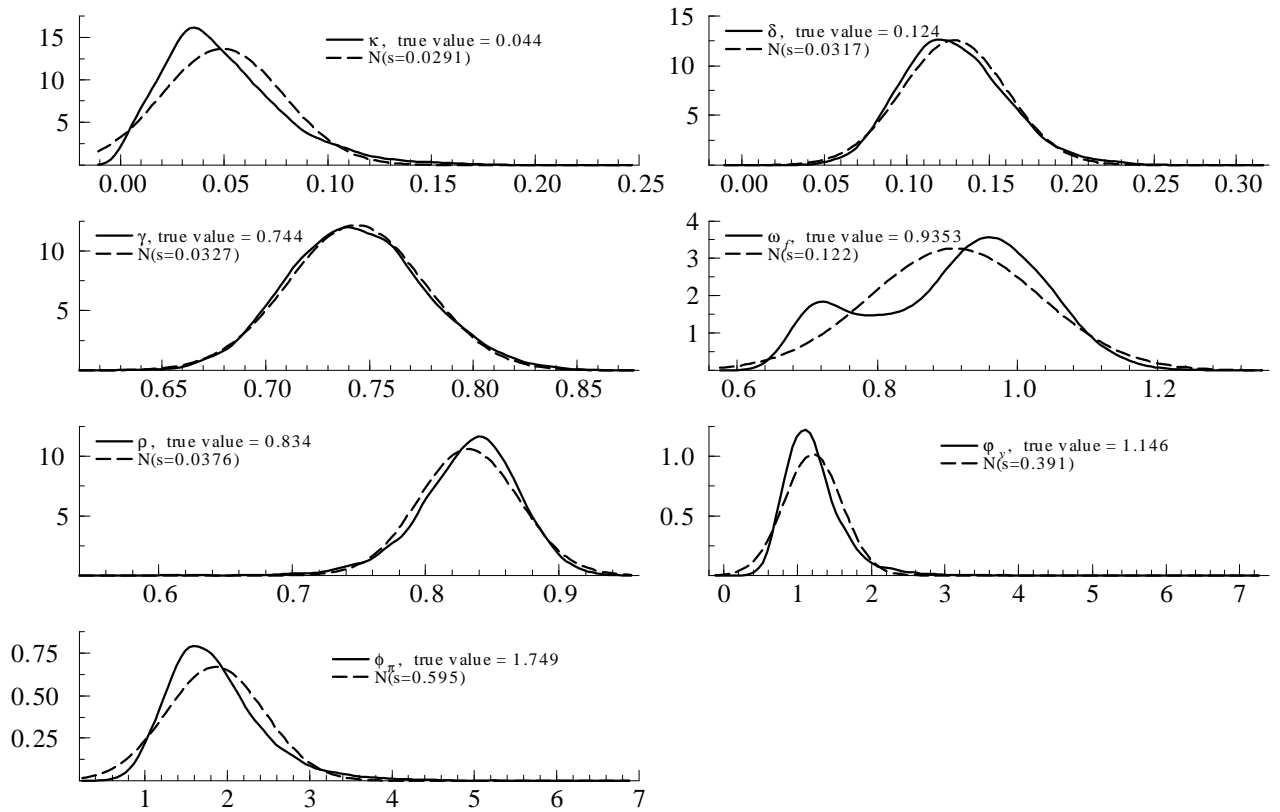
Figure 2: Empirical marginal distribution of the ML estimators of some parameters of the NK-DSGE model (solid lines) obtained from the Monte Carlo experiment against the Gaussian density (dotted lines) in samples of length $T = 500$.

List of Tables

Table 1: Monte Carlo results of $M = 10000$ replications of the size of tests of rank (LR1), long-run (LR2), and CER (LR3) of the NK-DSGE-model and averages across Monte Carlo simulations of the ML estimates of the structural parameters. Two lags are used in the VAR-representation.

Panel 1: Simulated empirical sizes $\hat{\psi}_i$ of tests of NK-DSGE model			
	Empirical size of tests		
Tests	$T = 100$	$T = 200$	$T = 500$
$LR1\,(\psi_1 = 0.01)$: test of rank	0.006	0.009	0.011
$LR2\,(\psi_2 = 0.02)$: test of beta	0.067	0.040	0.028
$LR3\,(\psi_3 = 0.02)$: test of CER	0.027	0.020	0.018
$LR1\left(\hat{\psi}_1\right) \to LR2\left(\hat{\psi}_2\right) \to LR3\left(\hat{\psi}_3\right)$: $\hat{\psi} = \sum_{i=1}^{3} \hat{\psi}_i$	0.100	0.069	0.057
Overall rejection frequency of DSGE model	0.098	0.067	0.056
Panel 2: Means and s.e. of ML estimates of structural parameters:			
$\hat{E}_{MC}\left(\hat{\theta}_i\right)$ $_{(s.e._{MC}(\hat{\theta}_i))}$			
Parameters	$T = 100$	$T = 200$	$T = 500$
$\kappa = 0.044$	0.095 (0.111)	0.067 (0.058)	0.049 (0.029)
$\delta = 0.12404$	0.149 (0.082)	0.136 (0.053)	0.129 (0.032)
$\gamma = 0.744$	0.740 (0.079)	0.745 (0.053)	0.744 (0.033)
$\omega_f = 0.93537$	0.955 (0.180)	0.932 (0.148)	0.912 (0.122)
$\rho = 0.834$	0.826 (0.085)	0.832 (0.060)	0.832 (0.037)
$\varphi_y = 1.146$	1.440 (1.197)	1.356 (0.841)	1.207 (0.390)
$\varphi_\pi = 1.749$	2.436 (1.680)	2.155 (1.206)	1.859 (0.592)
$\rho_y = 0.796$	0.768 (0.141)	0.784 (0.079)	0.789 (0.043)
$\rho_\pi = 0.418$	0.404 (0.205)	0.380 (0.079)	0.362 (0.157)
$\rho_i = 0.404$	0.394 (0.135)	0.402 (0.098)	0.404 (0.062)
$\sigma_{\tilde{y}}^2 = 0.055$	0.072 (0.041)	0.062 (0.024)	0.058 (0.013)
$\sigma_{\pi}^2 = 0.391$	0.450 (0.181)	0.429 (0.108)	0.421 (0.079)
$\sigma_{i}^2 = 0.492$	0.515 (0.164)	0.508 (0.106)	0.496 (0.053)
$\sigma_{y^p}^2 = 0.020$	0.020 (0.003)	0.020 (0.002)	0.020 (0.001)

Table 2: Unrestricted reduced form estimates of the $VAR(2)$ for $W_t^0:=(\tilde{y}_t, \pi_t, i_t)'$; U.S. quarterly data 1985q1-2008q3, T:=93 (+2 initial lags)

	$\tilde{y}_t$	π_t	i_t
$\tilde{y}_{t-1}$	1.041 (0.121)	0.045 (0.046)	0.075 (0.020)
π_{t-1}	-0.109 (0.297)	0.475 (0.113)	0.111 (0.049)
i_{t-1}	1.147 (0.516)	0.022 (0.196)	1.504 (0.085)
$\tilde{y}_{t-2}$	-0.135 (0.116)	-0.041 (0.044)	-0.072 (0.019)
π_{t-2}	-0.424 (0.288)	0.215 (0.110)	-0.018 (0.047)
i_{t-2}	-1.08 (0.493)	-0.021 (0.188)	-0.552 (0.081)
$const$	0.239 (0.190)	0.193 (0.072)	-0.035 (0.031)

$$\hat{\Sigma}_{W,\varepsilon}=\begin{pmatrix} 0.500 & -0.256 & 0.292 \\ & 0.190 & 0.117 \\ & & 0.082 \end{pmatrix}$$

Diagnostic tests			
LM AR 1-5 test:	2.93 [0.02]	2.73 [0.025]	0.99 [0.43]
LM ARCH 1-4 test:	0.20 [0.93]	0.87 [0.49]	0.77 [0.55]
LM Normality test:	0.27 [0.88]	7.08 [0.03]	2.93 [0.23]

LM AR 1-5 vector test: 1.67 [0.01]

LM vector Normality test:10.07 [0.12]

NOTES: The estimated unrestricted VAR covariance matrix $S_{W,e}$ is reported such that correlations appear in the off-diagonal terms. Asymptotic standard errors are reported in parentheses below estimates, P-values in brackets. 'LM AR 1-5 test' is the test for the absence of residuals autocorrelation against the alternative of correlation uo tp 5 lags; 'LM ARCH 1-4' test for the absence of ARCH components in the disturbances against the alternative of ARCH components up to lag order 4; 'LM Normality test' is the test for the null of Gaussian disturbances. Estimation is carried out by considering within-periods initial values.

Table 3: Estimated eigenvalues of the companion matrix associated with the unrestricted VAR(2) for for $W_t^0 := (\tilde{y}_t, \pi_t, i_t)'$ and LR "one shot" Trace test for cointegration rank $r := 2$ (one unit root) vs $r := 3 := p$ (stationary system); U.S. quarterly data 1985q1-2008q3, T:=93 (+2 initial lags).

Estimated roots		
real	imaginary	modulus
0.9091	0	0.9091
0.8189	0.1257	0.8295
0.8189	-0.1257	0.8295
0.7420	0	0.7420
-0.3202	0	0.3202
0.0495	0	0.0495
Test of rank		
LR for $r=2$ vs $r=3$ [LR1&2]		4.01 [0.045]

 NOTES: Estimation is carried out by considering within-periods initial values. P-values in brackets.

Table 4: ML estimates of the structural parameters θ of the NK-DSGE system (10)-(13) for for $W_t^0 := (\tilde{y}_t, \pi_t, i_t)'$; U.S. quarterly data 1985q1-2008q3, T:=93 (+2 initial lags).

Parameters in θ_s	Interpretation	$\mathrm{ML}_{\mathrm{det}}$
γ	AD, forward look. term	0.735 (0.092)
δ	AD, inverse elasticity of sub.	0.057 (0.013)
$\varkappa$	NKPC, indexation	0.021 (0.036)
implied value ω_f	NKPC, forward-looking	0.969 (0.034)
κ	NKPC, slope	0.058 (0.023)
ρ	Policy rule, smoothing term	0.623 (0.404)
$\varphi_{\tilde{y}}$	Policy rule, react. to out. gap	0.233 (0.680)
φ_π	Policy rule, react. to inflation	5.21 (3.18)
$\rho_{\tilde{y}}$	AD, disturbance persist.	0.889 (0.179)
ρ_π	NKPC, disturbance persist.	0.921 (0.347)
ρ_R	Policy rule, disturbance persist.	0.826 (0.515)

$$
\hat{\Sigma}_{W,u} = \begin{pmatrix}
\underset{(0.00029)}{0.0196} & \underset{(0.00031)}{-0.0018} & \underset{(0.0017)}{0.0014} \\
 & \underset{(0.00042)}{0.0029} & \underset{(0.0026)}{-0.014} \\
 & & \underset{(0.022)}{0.148}
\end{pmatrix}
$$

LR test for CER ($LR3$)
Constrained log-likelihood =60.855
Unrestricted log-likelihood =69.69
$LR3$=17.66 [0.024]

NOTES: ML estimates have been obtained from the determinate VAR(2) solution for $W_t^o := (\tilde{y}_t, \pi_t, i_t)'$ in Eq. (3) by maximizing the Gaussian log-likelihood under the CER in Eq.s (4)-(6). The variables in W_t^o have been preliminarily demeaned. The constrained VAR(2) log-likelihood function has been maximized by the BFGS method using a grid search for δ (range [0.05, 0.15]), ϱ (range [0.03, 0.06]), $\varkappa$ (range [0.02, 0.05]), φ_y (range [0.10, 1.50]) and φ_π (range [1, 5.5]), and estimating γ, ρ, $\rho_{\tilde{y}}$, ρ_π and ρ_R freely. Asymptotic standard errors are reported in parentheses below estimates. The LR test for the CER (LR3) has 24-16=8 degree of freedom.