cemmap

centre for microdata methods and practice

Model comparisons in unstable environments

Raffaella Giacomini
Barbara Rossi

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP13/12

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Model Comparisons in Unstable Environments

Raffaella Giacomini and Barbara Rossi

(UCL and CeMMAP) *(ICREA, UPF, CREI and Barcelona GSE)*

June 7, 2012

Abstract

The goal of this paper is to develop formal tests to evaluate the relative in-sample performance of two competing, misspecified non-nested models in the presence of possible data instability. Compared to previous approaches to model selection, which are based on measures of global performance, we focus on the local relative performance of the models. We propose three tests that are based on different measures of local performance and that correspond to different null and alternative hypotheses. The empirical application provides insights into the time variation in the performance of a representative DSGE model of the European economy relative to that of VARs.

Keywords: Model Selection Tests, Misspecification, Structural Change, Kullback-Leibler Information Criterion

Acknowledgments: We are grateful to M. del Negro, F. Smets, R. Wouters, W.B. Wu and Z. Zhao for sharing their codes. We also thank the editor, three anonymous referees, seminar participants at the Empirical Macro Study Group at Duke U., Atlanta Fed, UC Berkeley, UC Davis, U. of Michigan, NYU Stern, Boston U., U. of Montreal, UNC Chapel Hill, U. of Wisconsin, UCI, LSE, UCL, Stanford's 2006 SITE workshop, the 2006 Cleveland Fed workshop, the 2006 Triangle Econometrics workshop, the Fifth ECB Workshop, the 2006 Cass Business School Conference on Structural Breaks and Persistence, the 2007 NBER Summer Institute, the 2009 NBER-NSF Time Series Conference and the 2012 AEA Meetings for useful comments and suggestions. Support by NSF grants 0647627 and 0647770 is gratefully acknowledged.

J.E.L. Codes: C22, C52, C53

1 Introduction

The problem of detecting time-variation in the parameters of econometric models has been widely investigated for several decades, and empirical applications have documented that structural instability is widespread.

In this paper, we depart from the literature by focusing on investigating instability in the performance of models, rather than focusing solely on instability in their parameters. The idea is simple: in the presence of structural change, it is plausible that the performance of a model may itself be changing over time, even if the model's parameters remain constant. In particular, when the problem is that of comparing the performance of competing models, it would be useful to understand which model performed better at which point in time.

The goal of this paper is therefore to develop formal techniques for conducting inference about the relative performance of two models over time, and to propose tests that can detect time variation in relative performance even when the parameters are constant. Existing model selection tests such as Rivers and Vuong (2002) are inadequate for answering this question, since they work under the assumption that there exists a globally best model. The central idea of our method is instead to propose a measure of the models' local relative performance: the "local relative Kullback-Leibler Information Criterion" (local relative KLIC), which represents the relative distance of the two (misspecified) likelihoods from the true likelihood at a particular point in time. We then investigate ways to conduct inference about the local relative KLIC and construct tests of the joint null hypothesis that the relative performance and the parameters of the models are constant over time.

We propose three tests, which correspond to different assumptions about the parameters and the relative performance under the null and alternative hypotheses: 1) a "one-time reversal" test against a one-time change in models' performance and parameters; 2) a "nonparametric test" and 3) a "fluctuation test" against smooth changes in both performance and parameters. The first test is based on estimating the parameters and the relative performance before and after potential change dates, whereas the latter two are based on nonparametric estimates of local performance and local parameters. While the second and third tests consider the same test statistic, they differ in the asymptotic approximation that we use to derive its distribution under the null hypothesis (which also has a different formulation). The nonparametric test adopts the standard shrinking-bandwidth approximation of Wu and Zhao (2007), whereas the fluctuation test is based on a novel fixed-bandwidth approximation which we show delivers a better finite-sample performance.

For all three tests, we show that the dependence of the local performance on unobserved parameters does not affect the asymptotic distribution of the test statistic, as long as the parameters are also estimated locally. This can be viewed as an extension of a similar finding in Rivers and

Vuong (2002) to our local setting, given that the parameters are estimated by maximizing the same criterion on which the performance measure is based.

Our research is related to several papers in the literature, in particular Rossi (2005) and, more distantly, to Muller and Petalas (2009), Elliott and Muller (2005), Andrews and Ploberger (1994) and Andrews (1993). Rossi (2005) proposes a test that is similar to our one-time reversal test but focuses on the case of nested and correctly specified models. Here we consider the more general case of non-nested and misspecified models and propose two additional tests. In a companion paper, Giacomini and Rossi (2010) investigate the problem of estimating and testing the time variation in the relative performance of models in an out-of-sample forecasting context. Even though some of the techniques are similar, the additional complication in the in-sample context considered in this paper is that the measure of relative performance depends on estimated parameters, which need to be taken into account when performing inference. The dependence on parameter estimates can instead be ignored in an out-of-sample context, provided one adopts the asymptotic approximation with finite estimation window considered by Giacomini and Rossi (2010).

Our approach in this paper is also related to the literature on parameter instability testing (e.g., Brown, Durbin and Evans, 1975; Ploberger and Kramer, 1992; Andrews, 1993; Andrews and Ploberger, 1994; Elliott and Muller, 2005; Muller and Petalas, 2009) in that we adapt the tools developed in that literature to our different context of testing the joint hypothesis that the relative performance of the models is equal at each point in time and that the parameters are constant.

One important limitation of our approach is that our methods are not applicable when the competing models are nested, which is common in the literature on model selection testing based on Kullback-Leibler-type of measures. See Rivers and Vuong (2002) for an in-depth discussion of this issue.

The paper is structured as follows. The next section discusses a motivating example that illustrates the procedures proposed in this paper. Section 3 defines the tests. Section 4 evaluates the small sample properties of our proposed procedures in a Monte Carlo experiment, and Section 5 presents the empirical results. Section 6 concludes. The proofs are collected in the appendix.

2 Motivating Example

Let $y_t = \beta_t^0 x_t + \gamma_t^0 z_t + u_t$, with $u_t \sim i.i.d.N(0,1)$, x_t, z_t independent $N(0, \sigma_{x,t}^2)$ and $N(0, \sigma_{z,t}^2)$, respectively, independent of each other and of u_t for $t = 1, ..., T$, so that the true conditional density of y_t is $h_t : N(\beta_t^0 x_t + \gamma_t^0 z_t, 1)$. Suppose the researcher's goal is to compare two misspecified models: model 1, which specifies a density $f_t : N(\beta_t^0 x_t, 1)$ and model 2, with density $g_t : N(\gamma_t^0 z_t, 1)$. To measure the relative distance of f_t and g_t from h_t at time t we propose using the relative Kullback-Leibler Information Criterion at time t, $\Delta KLIC_t$, (henceforth the "local relative KLIC"), defined as:

$$\text{Local relative KLIC} : \Delta KLIC_t(\theta_t^0) = E\left[\ln\left(h_t/g_t\right)\right] - E\left[\ln\left(h_t/f_t\right)\right] = E\left[\ln f_t - \ln g_t\right], \tag{1}$$

where $\theta_t^0 = (\beta_t^{0\prime}, \gamma_t^{0\prime})'$ and the expectation is taken with respect to the true density h_t. If $\Delta KLIC_t(\theta_t^0) > 0$, model 1 performs better than model 2 at time t. In our example, it can be easily shown that[1]:

$$\Delta KLIC_t(\theta_t^0) = \frac{1}{2}\left[\left(\beta_t^0\right)^2 \sigma_{x,t}^2 - \left(\gamma_t^0\right)^2 \sigma_{z,t}^2\right]. \tag{2}$$

Intuitively, $\Delta KLIC_t(\theta_t^0)$ measures the relative degree of mis-specification of the two models at time t. For model 2, the contribution of its mis-specification is reflected in the contribution of the omitted variable x_t to the variance of the error term, which equals $\left(\beta_t^0\right)^2 \sigma_{x,t}^2$. Similarly, the mis-specification of model 1 is measured by $\left(\gamma_t^0\right)^2 \sigma_{z,t}^2$. Thus, model 2 performs better than model 1 if the contribution of its mis-specification to the variance of the error is smaller than for model 1.

Importantly, equation (2) shows that the time variation in the relative KLIC reflects the time variation in the relative mis-specification of the two models. In particular, the time variation in relative performance might be due to the fact that β_t^0, γ_t^0 change in ways that affect $\Delta KLIC_t$ differently over time, but it might also be caused by $\sigma_{x,t}^2$ and $\sigma_{z,t}^2$ changing in different ways over time while the parameters remain constant.

As can be seen from expression (2), the main challenge in estimating the local relative KLIC is its dependence on the unknown parameters at time t. Our goal is to construct tests of equal performance over time that take into account such dependence. We propose three different tests, which correspond to different assumptions about the behavior of the parameters and of the relative performance under the alternative hypothesis.

The first test ("one-time reversal test") assumes that under the alternative hypothesis there is a one-time change in relative performance as well as (at most) a one-time change in parameters at the same time. This corresponds to the following null and alternative hypotheses:

$$H_0^{OT} : \{\Delta KLIC_t\left(\theta_t\right) = 0\} \cap \{\theta_t = \theta\} \text{ for } t = 1, ..., T \tag{3}$$

where

$$\theta = (\beta', \gamma')' \text{ with } \beta = \arg\max_b E\left[\frac{1}{T}\sum_{t=1}^{T} \ln f_t(b)\right] \tag{4}$$

(and similarly for γ), and

$$\begin{aligned} H_1^{OT} \quad &: \quad \cup_{\delta\in\Pi}\{\Delta KLIC_t\left(\theta_t\right) = \mu_1\left(\delta\right) 1\left(t \leq [T\delta]\right) + \mu_2\left(\delta\right) 1\left(t > [T\delta]\right)\} \\ \cap\{\theta_t \quad &= \quad \theta_1\left(\delta\right) 1\left(t \leq [T\delta]\right) + \theta_2\left(\delta\right) 1\left(t > [T\delta]\right)\} \end{aligned} \tag{5}$$

[1] $\Delta KLIC_t = \frac{1}{2}E\left[\left(u_t - \beta_t^0 x_t\right)^2 - \left(u_t - \gamma_t^0 z_t\right)^2\right] = \frac{1}{2}E\left[\left(\beta_t^0\right)^2 x_t^2 - \left(\gamma_t^0\right)^2 z_t^2\right] = \frac{1}{2}((\beta_t^0)^2 \sigma_{x,t}^2 - \left(\gamma_t^0\right)^2 \sigma_{z,t}^2)$

for some $(\mu_1(\delta), \mu_2(\delta)) \neq (0,0)$, some $\delta \in \Pi \subset (0,1)$, $t = 1, ..., T$ and $\theta_t = (\beta_t, \gamma_t)'$, where $\beta_t = \beta_1(\delta) 1(t \leq [T\delta]) + \beta_2(\delta) 1(t > [T\delta])$, with

$$\beta_1(\delta) = \arg\max_b E\left[\frac{1}{[T\delta]} \sum_{t=1}^{[T\delta]} \ln f_t(b)\right], \tag{6}$$

$$\beta_2(\delta) = \arg\max_b E\left[\frac{1}{[T(1-\delta)]} \sum_{t=[T\delta]+1}^{T} \ln f_t(b)\right], \tag{7}$$

(and similarly for γ_t). Thus, β_t and γ_t are the local maximum likelihood "pseudo-true" parameters computed in the sub-samples before and after the reversal, which happens at the unknown fraction of the total sample δ.

The fact that the local parameter is maintained constant under the null hypothesis is not in principle necessary, but it makes the assumptions that underlie the validity of our test more plausible. We will further discuss this issue in Section 3.1 below. The approach focuses on the models' local relative performance by measuring it separately before and after the reversal. In case the null hypothesis is rejected, the time of the change δT can be estimated and the path of relative performance equals $\mu_1(\delta)$ before the change and $\mu_2(\delta)$ after the change.

The second and third tests ("nonparametric test" and "fluctuation test") involve estimating both the measure of relative performance $\Delta KLIC_t(\theta_t)$ and the parameters θ_t nonparametrically. The two tests are based on the same test statistic but consider two alternative asymptotic approximations and, as a result, correspond to different null and alternative hypotheses. The nonparametric test is based on the standard shrinking-bandwidth approximation adopted in the literature, where $\Delta KLIC_t(\theta_t)$ can be consistently estimated by kernel smoothing techniques. The test corresponds to the following null and alternative hypotheses:

$$H_0^{SB} : \{\Delta KLIC_t(\theta_t) = 0\} \cap \{\theta_t = \theta\} \text{ for } t = 1, ..., T \tag{8}$$

with θ as in (4) and

$$H_1^{SB} : \Delta KLIC_t(\theta_t) = \mu(t/T, \theta(t/T)) \neq 0 \text{ at some } 1 \leq t \leq T,$$

for some smooth functions $\mu(\cdot)$ and $\theta(\cdot)$.

A possible concern with the standard shrinking-bandwidth approximation is that it might perform poorly in small samples, such as those available to macroeconomists. We thus derive the fluctuation test using a novel asymptotic approximation where the bandwidth is fixed. In this approximation, consistent estimation of the local relative performance is not possible, but what can be consistently estimated is a different measure of relative performance, which is a smoothed version of the local relative KLIC:

$$\text{Smoothed local relative KLIC} : \Delta KLIC_t^*(\theta_t) = E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\left(\ln f_j(\beta_t) - \ln g_j(\gamma_t)\right)\right], \tag{9}$$

where $K(\cdot)$ is a kernel function, and h is the bandwidth.

The fluctuation test corresponds to different null and alternative hypotheses:

$$H_0^{FB} : \{\Delta KLIC_t^*(\theta_t) = 0\} \cap \{\theta_t = \theta\} \text{ for } t = m/2, ..., T - m/2, \tag{10}$$

with θ as in (4) and

$$H_1^{FB} : \Delta KLIC_t^*(\theta_t) \neq 0 \text{ at some } m/2 \leq t \leq T - m/2,$$

where $\theta_t = (\beta_t, \gamma_t)'$ and

$$\beta_t = \arg\max_b E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\ln f_j(\beta)\right], \tag{11}$$

(and similarly for γ_t). In particular, when using a rectangular kernel we have, under the alternative hypothesis:

$$\Delta KLIC_t^*(\theta_t) = E\left[\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left(\ln f_j(\beta_t) - \ln g_j(\gamma_t)\right)\right], \tag{12}$$

and β_t and γ_t are the local maximum likelihood pseudo-true parameters computed over the estimation window of length m, so that, e.g.,[2]

$$\beta_t = \arg\max_b E\left[\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\ln f_j(b)\right]. \tag{13}$$

As in the case of the null hypothesis (8) the constancy of pseudo-true parameters under the null hypothesis is a stronger requirement than necessary, but it makes the assumptions underlying our test more plausible. In the example, the smoothed local relative KLIC is

$$\Delta KLIC_t^*(\theta_t) = \frac{1}{2}\left[\begin{array}{c} \frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left(\beta_j^0\right)^2\sigma_{x,j}^2 - \frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left(\gamma_j^0\right)^2\sigma_{z,j}^2 \\ +\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left(\gamma_j^0-\gamma_t\right)^2\sigma_{z,j}^2 - \frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left(\beta_j^0-\beta_t\right)^2\sigma_{x,j}^2 \end{array}\right], \tag{14}$$

which is a different object than $\Delta KLIC(\theta_t^0)$ in (2), since it can be shown that, even in this simple example, $\beta_t \neq \beta_t^0$ (in particular, here we have $\beta_t = \left(\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\beta_j^0\right) / \left(\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\sigma_{x,j}^2\right)$).

[2] We use the terminology "local" MLE because these are MLE estimators obtained using a sub-sample of the available data.

The difference between the various alternative hypotheses as well as the difference between the relative and the smoothed relative KLIC is clarified by Figure 1, which shows an example of two different types of time variation in relative performance that could arise in the context of the simple example considered in this section. In the first scenario (left panels of Figure 1) the time variation in relative performance is due to β_t^0 varying smoothly as a random walk whereas γ_t^0, $\sigma_{x,t}^2$, $\sigma_{z,t}^2$ are constant, $t = 1, ..., 100$. In the second scenario (right panels of Figure 1), $\beta_t^0, \gamma_t^0, \sigma_{z,t}^2$ are constant but the relative performance is time-varying because $\sigma_{x,t}^2$ has a break at $T/2$.

INSERT FIGURE 1 HERE

Figures 1(a) and 1(b) report the local relative KLIC in equation (1) in the two scenarios, which is the object of interest in the shrinking-bandwidth approximation. Figures 1(c) and 1(d) show the local relative KLIC as well as the smoothed local relative KLIC computed using a bandwidth $m/T = 1/5$, which is the measure of relative performance in the fixed-bandwidth approximation. Note that Figures 1(a-d) report population quantities (that is, they assume that the parameters and variances are known). Finally, Figures 1(e) and 1(f) show the measure of relative performance that arises as a result of testing (10) and (3). One can see that all three measures of relative performance that we propose capture the time variation in the relative performance of the models over time.

In contrast, the large dot reported in panels (a-d) of Figure 1 shows the global relative KLIC ($T^{-1} \sum_{t=1}^{T} \Delta KLIC_t$), which compares the average performance of the models over the whole sample and which is the object of interest of existing tests in the literature (e.g., Rivers and Vuong (2002)). One can see that the global relative KLIC is very close to zero, which means that the Rivers and Vuong's (2002) test would not reject the null hypothesis that the models perform equally well. This occurs because in our example there are reversals in the relative performance of the models during the time period considered. Since model 1 is better than model 2 in the first part of the sample, but model 2 is better than model 1 in the second part of the sample by a similar magnitude, on average over the full sample the two models have similar performance. However, the figure shows that the relative performance did change over time, and that the existing approaches would miss this important feature of the data, whereas our approach would be able to reveal which model performed best at different points in time.

In the following section, we develop the theory for the three statistical tests. The one-time reversal test of hypothesis (3) can be intuitively viewed as performing a Rivers and Vuong's (2002) test of equal performance allowing for one structural break under the alternative. The nonparametric test of hypothesis (8) relies on constructing simultaneous confidence bands for the local relative KLIC in (1) under the null hypothesis by adapting the shrinking-bandwidth approximation of Wu and Zhao (2007) to our different context. Finally, the fluctuation test of hypothesis (10) relies on constructing simultaneous confidence bands for a different object - the smoothed local relative KLIC in (9) - under the null hypothesis by using an alternative fixed-bandwidth approximation. We refer to this test as the fluctuation test in analogy with the literature on parameter stability testing (Brown et al. 1975 and Ploberger and Kramer 1992). Even though one can see that our tests draw on the existing literature on parameter instability testing, we face additional challenges in particular due to the fact that we are testing joint hypotheses of equal performance and stability and that the measure of performance depends on unknown local parameters.

The three tests involve trade-offs, some of which are highlighted by Figure 1. The first consideration is what type of alternative hypothesis seems more appropriate in a given situation. If the type of variation under the alternative hypothesis is a one-time change, the nonparametric test based on the local relative of Figure 1(b) and the one-time reversal test (Figure 1(f)) will accurately capture it, whereas the fluctuation test (which relies on the smoothed local relative KLIC of Figure 1(d)) will smooth out the time variation and thus make it more difficult to detect, implying a power loss for the test. This is also the case when one postulates a smooth change under the alternative hypothesis, in which case the fluctuation test (Figure 1(c)) should have lower power than the other tests because of its smoothing out of the time variation. The one-time reversal test would also be suboptimal in this context because it is based on an approximate measure of time variation, as can be seen in Figure 1(e). The previous discussion may lead one to think that the nonparametric test dominates the other two. All these considerations are however based on the asymptotic power of the test. In finite samples, instead, there is a concern that the asymptotic approximation which underlies the nonparametric test may perform poorly in finite samples. We investigate this possibility in the Monte Carlo section below and conclude that this concern is indeed a real one and thus end up not recommending the nonparametric test, at least for samples of the sizes typically available in macroeconomic applications.

How would the tests that we propose be implemented in practice? We provide an example in Figures 1(e-h). For the fluctuation test we provide boundary lines that would contain the time path of the smoothed local relative KLIC with a pre-specified probability level under the null hypothesis that the relative performance of the models is equal. Figures 1(e,f) depict such boundary lines. Clearly, the test rejects the hypothesis that the relative performance is the same. When this happens, researchers can rely on visual inspection of the (estimated) smoothed local relative KLIC to ascertain which model performed best at any point in time.

Figures 1(g,h) illustrate the one-time reversal test[3] for the two cases. The procedure estimates the time of the largest change in the relative performance, and then fits measures of average performance separately before and after the reversal. Figure 1(h) shows that when the true underlying relative performance has a sharp reversal, such as in the second scenario, then the procedure will

[3]The One-time Reversal test is implemented as a Sup-type test. See Section 3.1 for more details.

accurately estimate its time path. However, when the true underlying relative performance evolves smoothly over time, then the procedure will approximate it with a sharp reversal, as depicted in Figure 1(g). In both cases, the one-time reversal test strongly rejects the null hypothesis of equal performance.

3 Tests of Stability in the Relative Performance of Models

In this section, we consider the problem of conducting inference about the local relative performance of two models. In particular, we will propose three types of statistical tests of the hypothesis that the models have equal performance at each point in time. The tests differ in the measure of relative performance (e.g., the local relative KLIC for the nonparametric and one-time reversal tests vs. the smoothed local relative KLIC for the fluctuation test), the null and alternative hypotheses considered (e.g., smooth time variation under the alternative for the nonparametric and fluctuation tests vs. one-time change for the one-time reversal test), and the asymptotic approximation adopted in deriving the distribution of the test (shrinking-bandwidth for the nonparametric test vs. fixed-bandwidth for the fluctuation test).

In all the following sections, we assume that the user has available two possibly misspecified parametric models for the variable of interest y_t. The models can be multivariate, dynamic and nonlinear. In line with the literature (e.g., Vuong (1989) and Rivers and Vuong (2002)), an important restriction is that the models must be non-nested, which loosely speaking means that the models' likelihoods cannot be obtained from each other by imposing parameter restrictions.

3.1 The One-time Reversal Test

The object of interest for the one-time reversal test is the local relative KLIC, which measures relative performance as the relative distance of the two models from the true, unknown, data-generating process at time t :

$$\begin{aligned} \Delta KLIC_t(\theta_t) &= E[\Delta L_t(\theta_t)] = E[\ln f_t(\beta_t) - \ln g_t(\gamma_t)], \\ \text{for } t &= 1, ..., T, \end{aligned} \tag{15}$$

where f_t and g_t are the likelihoods for the two models and $\theta_t = (\beta_t', \gamma_t')'$ are such that:

$$\beta_t = \arg\max_{\beta \in B} E[\ln f_t(\beta)],$$

B a compact parameter space. A similar definition holds for γ_t, which depends on $g_t(\gamma)$.

This section derives tests that are designed for a specific form of time variation in the relative performance of the models under the alternative hypothesis, namely a one-time reversal in the relative performance and in the parameters, which occur at the same time. Let us define the time path of the relative performance under time variation as follows:

$$\Delta KLIC_t(\theta_t) = \mu_1(\delta) \cdot 1(t \leq [T\delta]) + \mu_2(\delta) \cdot 1(t > [T\delta]), \ t = 1, 2, .., T,$$

where $\theta_t = \theta(\delta) = \theta_1(\delta) \cdot 1(t \leq [T\delta]) + \theta_2(\delta) \cdot 1(t > [T\delta])$, δ denotes the time of the reversal as a fraction of the sample size, $\theta_1(\delta) = (\beta_1(\delta)', \gamma_1(\delta)')'$, $\theta_2(\delta) = (\beta_2(\delta)', \gamma_2(\delta)')'$ with

$$\begin{aligned}
\beta_1(\delta) &= \arg\max_\beta E\left[\left(\frac{1}{[T\delta]}\sum_{t=1}^{[T\delta]} \ln f_t(\beta)\right)\right], \\
\beta_2(\delta) &= \arg\max_\beta E\left[\left(\frac{1}{[T(1-\delta)]}\sum_{t=[T\delta]+1}^{T} \ln f_t(\beta)\right)\right],
\end{aligned}$$

(and similarly for $\gamma_1(\delta)$ and $\gamma_2(\delta)$).

Consider the problem of testing

$$\begin{aligned}
H_0^{OT} &: \{\Delta KLIC_t(\theta_t) = 0\} \cap \{\theta_t = \theta\} \text{ for } t = 1, ..., T && (16)\\
i.e.,\ H_0^{OT} &: \mu_t(\theta) \equiv \Delta KLIC_t(\theta) = 0 \text{ for } t = 1, ..., T, && (17)
\end{aligned}$$

where $\theta = [\beta', \gamma']'$, versus the alternative

$$\begin{aligned}
H_1^{OT} &: \cup_{\delta \in \Pi}\{\Delta KLIC_t(\theta_t) = \mu_1(\delta) 1(t \leq [T\delta]) + \mu_2(\delta) 1(t > [T\delta])\} && (18)\\
\cap \{\theta_t &= \theta_1(\delta) 1(t \leq [T\delta]) + \theta_2(\delta) 1(t > [T\delta])\} \\
\text{for some } (\mu_1(\delta), \mu_2(\delta)) &\neq (0,0), \text{ some } \delta \in \Pi \subset (0,1), \ t = 1, ..., T.
\end{aligned}$$

Note that the null hypothesis of interest is a possibly non-linear restriction on the parameters.

Given δ, the local maximum likelihood estimator $\widehat{\mu}(\delta) \equiv [\widehat{\mu}_1(\delta), \widehat{\mu}_2(\delta)]$ is given by

$$\widehat{\mu}_1(\delta) = \frac{1}{[T\delta]}\sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\theta}_1(\delta)\right), \ \widehat{\mu}_2(\delta) = \frac{1}{[T(1-\delta)]}\sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\theta}_2(\delta)\right) \qquad (19)$$

where $\widehat{\theta}_1(\delta) = (\widehat{\beta}_1(\delta)', \widehat{\gamma}_1(\delta)')'$, $\widehat{\theta}_2(\delta) = (\widehat{\beta}_2(\delta)', \widehat{\gamma}_2(\delta)')'$ with

$$\begin{aligned}
\widehat{\beta}_1(\delta) &= \arg\max_\beta \left(\frac{1}{[T\delta]}\sum_{t=1}^{T} \ln f_t(\beta) \cdot 1(t \leq [T\delta])\right) \\
\widehat{\beta}_2(\delta) &= \arg\max_\beta \left(\frac{1}{[T(1-\delta)]}\sum_{t=1}^{T} \ln f_t(\beta) \cdot 1(t > [T\delta])\right),
\end{aligned}$$

(and similarly for $\widehat{\gamma}(\delta)$). Also, let $\widehat{\beta}_T(\delta) = \arg\max_\beta \left(\frac{1}{T}\sum_{t=1}^{T} \ln f_t(\beta)\right)$ (and similarly for $\widehat{\gamma}_T(\delta)$), and $\widehat{\theta}_T(\delta) \equiv \left[\widehat{\beta}_T(\delta)', \widehat{\gamma}_T(\delta)'\right]'$.

Let $\Delta L_t(\theta(\delta))$ be distributed according to a parametric density whose likelihood be denoted by $\xi_T(\mu_t(\theta), \delta)$ (the latter is a function of the data, although we do not make the dependence explicit to simplify notation). Also, let $Q_\delta(.)$ denote a weight function that, for each δ, gives the same weight to ellipses associated with Wald-type tests of the null hypothesis (16) for the case in which δ is fixed and known. Let $J(\delta)$ be an integrable weight function on the values of δ. The Likelihood Ratio (LR) statistic for testing the null hypothesis (16), which implies $\xi_T(0) \equiv \xi_T(0,0)$, against a local alternative of the form $\xi_T\left(\mu^* T^{-1/2}, \delta\right)$ for some $\mu^* \equiv [\mu_1^*, \mu_2^*]'$ is:

$$LR_T = \frac{\int \xi_T\left(\mu^* T^{-1/2}, \delta\right) dQ_\delta(\mu^*)\, dJ(\delta)}{\int \xi_T(0)\, dQ_\delta(\mu^*)\, dJ(\delta)}, \tag{20}$$

By the Neyman-Pearson Lemma, a test based on LR_T is a best test for a given significance level for testing the simple null hypothesis that $\xi_T(0)$ is the true density versus the simple alternative that $\int \xi_T\left(\mu^* T^{-1/2}, \delta\right) dQ_\delta(\mu^*)\, dJ(\delta)$ is true, and has the best weighted average power for testing the simple null that $\xi_T(0)$ is the true density versus the alternative that $\xi_T\left(\mu^* T^{-1/2}, \delta\right)$ is the true density for some $\mu^* \in \mathbb{R}^2$, $\delta \in \Pi$.[4] Note that the weighted average power is constructed against the alternatives for $\Delta KLIC_t$, not in terms of the actual parameters, since the researchers' main interest is on $\Delta KLIC_t$.

Theorem 1 shows that the LR_T test statistic is asymptotically equivalent to an exponential-Wald test derived as follows. Let $\mathcal{I}_{0,\delta} \equiv -E\left[T^{-1}\frac{\partial^2}{\partial\mu\partial\mu'}\xi_T(\mu,\delta)\right]$, $H \equiv \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, and $\mathcal{I}_{T,\delta}$ be a consistent estimator for $\mathcal{I}_{0,\delta}$.

Assumptions OT: 1 $\left\{T^{-1/2}\sum_{j=1}^{[\tau T]}\Delta L_j(\theta)\right\}$ *obeys a Functional Central Limit Theorem (FCLT),* $\theta_t \in \Theta$, Θ *compact; (2) there exists a positive definite matrix* V_δ *such that* $V_\delta^{-1/2}\sqrt{T}\left(\widehat{\beta}(\delta) - \beta\right) \xrightarrow{d} N(0, I)$, *as* $T \to \infty$ *(and similarly for* $\widehat{\gamma}(\delta)$, $\widehat{\beta}_T(\delta)$, $\widehat{\gamma}_T(\delta)$*), and* β, γ *are interior to the parameter space*; *(3)* $T^{-1}\sum_{t=1}^{[T\delta]}\nabla \ln f_t(\beta)$ *satisfies a Uniform Law of Large Numbers* $\forall\delta \in \Pi$ *and* $T^{-1}\sum_{t=1}^{T}\nabla \ln f_t(\beta)$ *satisfies the Law of Large Numbers (and similarly for* $\nabla \ln g_t(\gamma)$*); (4) Under* H_0 , $E(\Delta L_t(\beta,\gamma)) = 0$ *and the distribution of* $\Delta L_t(\beta,\gamma)$ *does not depend on* δ $\forall\theta$ *and* $\Delta L_t(\beta,\gamma)$ *satisfying the null hypothesis; (5)* $Q_\delta(.) = N\left(0, c\mathcal{I}_{0,\delta}^{-1}\right)$ *for every* $\delta \in \Pi$ *and for some constant* $c > 0$ *and* $J(.)$ *is the uniform distribution; (6)* $sup_{\delta\in\Pi}||\widehat{\mu}(\delta)|| \underset{p}{\to} 0$ *and* $sup_{\delta\in\Pi}||\widehat{\theta}(\delta) - \theta(\delta)|| \underset{p}{\to} 0$ *under* H_0.[5]

Assumption OT(1) assumes a FCLT for partial sum processes. Assumptions OT(2,3) are standard ML assumptions that guarantee that the estimated parameters in our object of interest as well as the score functions obey regularity conditions ensuring their convergence. Assumption OT(4) specifies the null hypothesis. Assumption OT(5) specifies the weight function over the local alternatives; in practice, we will let $\Pi = \{0.15, ..., 0.85\}$. Assumption OT(6) assumes that the model is

[4] *Note that* $\int \xi_T(0)\, dQ_\delta(\mu^*)\, dJ(\delta) = \xi_T(0)$.

[5] See Andrews (1993, Lemma A-1) for primitive conditions ensuring Assumption OT(6).

sufficiently regular so that the estimators are consistent under the null hypothesis uniformly over $\delta \in \Pi$. Under these Assumptions we derive the following theorem:

Theorem 1 (One-Time Test) *Define the Exponential Wald test, $ExpW_T^*$, as:*

$$\begin{aligned} W_T(\delta) &= T\widehat{\mu}(\delta)' H' \left(H\mathcal{I}_{T,\delta}^{-1}H'\right)^{-1} H\widehat{\mu}(\delta) \\ ExpW_T^* &= (1+c)^{-1/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c}W_T(\delta)\right) dJ(\delta) \end{aligned} \quad (21)$$

Under Assumption OT: (i) Under H_0 described by (16), $LR_T - ExpW_T^* \underset{p}{\rightarrow} 0$. (ii) Under the local alternatives in (20), (21) is the test with the greatest weighted average power for the weight functions described in Assumption OT(5).

Note that Assumption OT(1) implies that $\left\{T^{-1/2}\sum_{t=1}^{[\delta T]} \Delta L_t(\theta)\right\} \underset{d}{\rightarrow} N(0, \delta\Psi)$ and $\left\{T^{-1/2}\sum_{t=[\delta T]}^{T} \Delta L_t(\theta)\right\} \underset{d}{\rightarrow} N(0, (1-\delta)\Psi)$ for any given δ, where $\Psi \equiv \lim_{T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} \Delta L_t(\theta)\right)$. Thus, asymptotically, a consistent estimate of $I_{0,\delta}$ both under H_0^{OT} and under H_1^{OT} is $I_{T,\delta}$, where:

$$\mathcal{I}_{T,\delta} = -\begin{pmatrix} \delta\widehat{\Sigma}_1 & 0 \\ 0 & (1-\delta)\widehat{\Sigma}_2 \end{pmatrix}, \quad (22)$$

$$\widehat{\Sigma}_1 = \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|) \frac{1}{[T\delta]} \sum_{j=1}^{[T\delta]} \left[\Delta L_{1,j}^d\left(\widehat{\theta}_1(\delta)\right)\right]^2, \quad (23)$$

$$\widehat{\Sigma}_2 = \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|) \frac{1}{[T(1-\delta)]} \sum_{j=[T\delta]+1}^{T} \left[\Delta L_{2,j}^d\left(\widehat{\theta}_2(\delta)\right)\right]^2 \quad (24)$$

$$\widehat{\sigma}^2 \equiv H\mathcal{I}_{T,\delta}^{-1}H' \quad (25)$$

$$\Delta L_{1,j}^d\left(\widehat{\theta}_1(\delta)\right) = \Delta L_j\left(\widehat{\theta}_1(\delta)\right) - \frac{1}{[T\delta]}\sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\theta}_1(\delta)\right) \quad (26)$$

$$\Delta L_{2,j}^d\left(\widehat{\theta}_2(\delta)\right) = \Delta L_j\left(\widehat{\theta}_2(\delta)\right) - \frac{1}{T-[T\delta]}\sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\theta}_2(\delta)\right). \quad (27)$$

The results of Theorem 1 hold in the presence of serial correlation as well as a one-time break in the variance at time $[\delta T]$ provided a heteroskedasticity and autocorrelation consistent estimator for the variance is used: cfr. Andrews and Ploberger (1994). The power properties of the test depend on c. Corollary 2 focuses on the limiting case where $c = 0$ and $c = \infty$, and their power properties will be evaluated in Section 4.

Corollary 2 *Suppose Assumption OT holds. Consider the test statistics*

$$\begin{aligned} ExpW^*_{\infty,T} &= \ln \frac{1}{1-2\delta_0} \int_{\delta_0}^{1-\delta_0} \exp\left(\frac{1}{2} W_T(\delta)\right) d\delta; \\ MeanW^*_T &= \frac{1}{1-2\delta_0} \int_{\delta_0}^{1-\delta_0} W_T(\delta)\, d\delta, \text{ where} \end{aligned}$$

$$W_T(\delta) = T \widehat{\mu}(\delta)' H' \left(H \mathcal{I}_{T,\delta}^{-1} H\right)^{-1} H \widehat{\mu}(\delta),$$

where $\delta_0 = 0.15$, $\widehat{\mu}(\delta)$ *is defined as in (19),* $\mathcal{I}_{T,\delta}$ *is as in (22). Under the null hypothesis (17),*

$$ExpW^*_{\infty,T} \implies \ln \frac{1}{1-2\delta_0} \int_{\delta_0}^{1-\delta_0} \exp\left(\frac{1}{2} \frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)} + \frac{1}{2}\mathcal{B}(1)^2\right) d\delta, \tag{28}$$

$$MeanW^*_T \implies \frac{1}{1-2\delta_0} \int_{\delta_0}^{1-\delta_0} \left(\frac{1}{2} \frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)} + \frac{1}{2}\mathcal{B}(1)^2\right) d\delta, \tag{29}$$

where $t = [\delta T]$ *and* $\mathcal{B}(\cdot)$ *and* $\mathcal{BB}(\cdot)$ *are, respectively, a standard univariate Brownian motion and a Brownian bridge, where* $\mathcal{BB}(\delta) \equiv \mathcal{B}(\delta) - \delta\mathcal{B}(1)$*. The null hypothesis is rejected when* $ExpW^*_{\infty,T} > \kappa_\alpha$ *and* $MeanW^*_T > \nu_\alpha$*. Simulated values of* $(\alpha; \kappa_\alpha, \upsilon_\alpha)$ *are:* $(0.05; 3.13, 5.36)$ *and* $(0.10; 2.44, 4.26)$.

We also provide Sup-type tests for the one-time reversal in the following proposition:[6]

Proposition 3 (Sup-type Test) *Suppose Assumption OT holds. Let* $QLR^*_T = \sup_{\delta\in\Pi} \Phi_T(\delta)$, $\Phi_T(\delta) = LM_1 + LM_2(\delta)$, *where*

$$\begin{aligned} LM_1 &= \hat{\sigma}^{-2} T^{-1/2} \left[\sum_{t=1}^{T} \Delta L_t\left(\widehat{\theta}_T\right)\right]^2 \\ LM_2(\delta) &= \hat{\sigma}^{-2} \frac{1}{\delta(1-\delta)} T^{-1/2} \left[(1-\delta) \sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\theta}_1(\delta)\right) - \delta \sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\theta}_2(\delta)\right)\right]^2, \end{aligned}$$

$\widehat{\sigma}^2$ *a consistent estimator of the asymptotic variance* $\sigma^2 = var\left(T^{-1/2} \sum_{t=1}^{T} \Delta L_t(\theta_t)\right)$, *for example (25). Under the null hypothesis (17), we have:* $QLR^*_T \Longrightarrow \sup_{\delta\in\Pi} \left[\frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)} + \mathcal{B}(1)^2\right]$, *where* $t = [\delta T]$, *and* $\mathcal{B}(\cdot)$ *and* $\mathcal{BB}(\delta) \equiv \mathcal{B}(\delta) - \delta\mathcal{B}(\delta)$ *are, respectively, a standard univariate Brownian motion and Brownian bridge. The null hypothesis is thus rejected when* $QLR^*_T > k_\alpha$*. The critical values* (α, k_α) *are:* $(0.05, 9.8257)$, $(0.10, 8.1379)$.

Among the advantages of this approach, we have that: (i) when the null hypothesis is rejected, it is possible to evaluate whether the rejection is due to instabilities in the relative performance or to a model being constantly better than its competitor; (ii) if such instability is found, it is possible

[6] Sup-type tests have been used in the parameter instability literature since Andrews (1993).

to estimate the time of the switch in the relative performance; (iii) the test is optimal against one time breaks in the relative performance. Here below is a step by step procedure to implement the approach suggested in Proposition 3 with an overall significance level α:

(i) test the hypothesis of equal performance at each time by using the statistic QLR_T^* from Proposition 3 at α significance level;

(ii) if the null is rejected, compare LM_1 and $\sup_{\delta\in\Pi} LM_2(\delta)$, with the following critical values: $(3.84, 8.85)$ for $\alpha = 0.05$, $(2.71, 7.17)$ for $\alpha = 0.10$, and $(6.63, 12.35)$ for $a = 0.01$. If only LM_1 rejects then there is evidence in favor of the hypothesis that one model is constantly better than its competitor. If only $\sup_{\delta\in\Pi} LM_2(\delta)$ rejects, then there is evidence that there are instabilities in the relative performance of the two models but neither is constantly better over the full sample. Note that the latter corresponds to Andrews' (1993) Sup-test for structural break. If both reject then it is not possible to attribute the rejection to a unique source.[7]

(iii) estimate the time of the reversal by $t^* = T \times \arg\sup_{\delta\in\{[0.15],...[0.85]\}} LM_2(\delta)$ and let $\delta^* \equiv [t^*/T]$.

(iv) to extract information on which model to choose, we suggest to plot the time path of the underlying relative performance as:

$$\begin{cases} \frac{1}{t^*}\sum_{t=1}^{t^*}\left(\ln f_t(\widehat{\beta}_1(\delta^*)) - \ln g_t(\widehat{\gamma}_1(\delta^*))\right) & \text{for } t \leq t^* \\ \frac{1}{(T-t^*)}\sum_{t=t^*+1}^{T}\left(\ln f_t(\widehat{\beta}_2(\delta^*)) - \ln g_t(\widehat{\gamma}_2(\delta^*))\right) & \text{for } t > t^* \end{cases}$$

3.2 The Nonparametric Test

The object of interest is again the local relative KLIC in equation (15). We consider the following null and alternative hypotheses:

$$\begin{aligned} H_0^{SB} &: \{\Delta KLIC_t(\theta_t) = 0\} \cap \{\theta_t = \theta\} \text{ for } t = 1, ..., T, \quad (30)\\ H_1^{SB} &: \Delta KLIC_t(\theta_t) = \mu(t/T, \theta(t/T)) \neq 0 \text{ at some } 1 \leq t \leq T, \text{ where } \mu(\cdot) \text{ and } \theta(\cdot) \in C^3[0,1]. \end{aligned}$$

The test relies on first constructing a nonparametric estimate of the local relative KLIC:

$$\Delta\widehat{KLIC}_t = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right)\Delta L_t\left(\widehat{\theta}(\tau)\right) \quad (31)$$

where $K(\cdot)$ is a kernel with $\int K(u)\,du = 1$, h is the bandwidth, $\Delta L_t(\theta) = \ln f_t(\beta) - \ln g_t(\gamma)$ and $\tau \in [0,1]$ is such that $t = [\tau T]$. We assume that the parameters of the models are also estimated

[7] This procedure is justified by the fact that the two components LM_1 and LM_2 are asymptotically independent – see Rossi (2005). Performing two separate tests does not result in an optimal test, but it is nevertheless useful to heuristically disentangle the causes of rejection of equal performance. The critical values for LM_1 are from a χ_1^2 whereas those for LM_2 are from Andrews (1993).

locally. E.g., the estimator $\widehat{\beta}(\tau)$ for the first model is the solution to

$$\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right) \nabla \ln f_t\left(\widehat{\beta}(\tau)\right) = 0, \tag{32}$$

where $\nabla \ln f_t(.)$ denotes the first derivative of the log-likelihood at time t. Note that, in the case of the rectangular kernel considered in Corollary 5 below, this in practice amounts to estimating the parameters of the models by local maximum likelihood over rolling windows of length Th.

A test of (30) can be obtained by deriving simultaneous confidence bands for the local relative KLIC by building on the framework of Wu and Zhao (2007). The test relies on the following assumptions:

Assumption SB: (1) $\Delta L_t(\theta_t) = \mu(t/T, \theta(t/T)) + \varepsilon_t$, $t = 1, ..., T$, *with* ε_t *such that* $max_{t \leq T} |S_t - \sigma \mathcal{B}(t)| = o_{AS}\left\{T^{1/4} \ln(T)\right\}$, *where* $S_t = \sum_{i=1}^{t} \varepsilon_i$, $\sigma^2 = \sum_{t=-\infty}^{\infty} E[\varepsilon_0 \varepsilon_t] > 0$ *and* $\mathcal{B}(.)$ *is a standard Brownian motion; (2)* $K(\cdot)$ *is a symmetric kernel with support* $[-w, w]$ *which belongs to the class* $H(\alpha)$ *as in Definition 1 of Wu and Zhao (2007); (3) The bandwidth* h *satisfies the condition* $Th \to \infty$, $h \to 0$, $\frac{\ln(T)^3}{h\sqrt{T}} + Th^7 \ln(T) \to 0$ *and* $\frac{\sqrt{Th}}{\ln(T)^3} \to \infty$*; (4)* $\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right)\left[\Delta s_t(\theta(\tau)) - E(\Delta s_t(\theta(\tau)))\right] = O_{as}(1)$ *uniformly in* $\theta(\tau)$ and τ, *where* $\Delta s_t(\theta) \equiv \partial \Delta L_t(\theta)/\partial \theta$*; (5) there exists a bias-adjusted local maximum likelihood estimator,* $\widetilde{\theta}(\tau)$, *such that, for every* τ, $\sqrt{Th}\left(\widetilde{\theta}(\tau) - \theta(\tau)\right) = O_{as}(1)$ *and* $\theta \in \Theta$, Θ *compact.*

Assumption SB is similar to the assumptions in Wu and Zhao (2007). The difference between our framework and theirs is that $\Delta L_t(\theta_t)$ is unknown in our case and thus needs to be estimated. Assumption SB(1), in particular, deserves further discussion. Even though it is possible to find primitive conditions for this strong invariance principle allowing for the error process ε_t to be dependent and stationary (as in Wu and Zhao, 2007), the assumption of stationarity for ε_t may be problematic in our context because of the dependence of the likelihood differences $\Delta L_t(\theta_t)$ on θ_t: it essentially amounts to assuming that the possible time variation in the parameters only affects the mean of the likelihood differences but not their higher moments. The assumption is however satisfied under the joint null hypothesis that the models have equal performance and that the parameters are constant, which is the reason why we impose constant parameters under the null hypothesis. The assumption that $\sigma > 0$ rules out the possibility that the models are nested (see the related discussion in Rivers and Vuong, 2002). Assumption SB(3) is the standard shrinking-bandwidth assumption made in the nonparametric literature, which guarantees that the local performance and the local parameters can be consistently estimated. Primitive conditions for Assumptions SB(4)-(5) can be derived in the context of specific examples.

Going back to the example in Section 2, we can see that the assumptions are satisfied since we have $\varepsilon_t = \Delta L_t - E[\Delta L_t] = \frac{1}{2}[\beta_t^2\left(x_t^2 - 1\right) - \gamma_t^2\left(z_t^2 - 1\right)] + \left(\beta_t x_t - \gamma_t z_t\right) u_t$, which has variance $\sigma_t^2 = \frac{1}{2}[\beta_t^4 + \gamma_t^4] + \beta_t^2 + \gamma_t^2$. Under the null hypothesis, $\beta_t = \beta$ and $\gamma_t = \gamma$ and thus ε_t is i.i.d. and σ_t^2 is constant, which imply that assumption SB is satisfied.

The following proposition gives the confidence bands which are the basis for the nonparametric test.

Proposition 4 *Under Assumption SB, asymptotic* $100(1-\alpha)\%$ *simultaneous confidence bands for* μ *are given by*

$$\widehat{\mu}(\tau) - h^2\Psi\widehat{\mu}(\tau)'' \pm \frac{\kappa\,\widehat{\sigma}}{\sqrt{Th}}\left[B_K - \frac{\ln\left[\ln(1-\alpha)^{-1/2}\right]}{\sqrt{2\ln\left(\frac{1}{h}\right)}}\right], \tag{33}$$

$$\Psi = \int K(u)u^2du/2;\ \kappa^2 = \int K^2(u)\,du \tag{34}$$

$$B_K = \sqrt{2\ln\left(\frac{1}{h}\right)} + \frac{1}{\sqrt{2\ln\left(\frac{1}{h}\right)}}\left[\frac{2-\psi}{2\psi}\ln(\ln(h^{-1})) + \ln\left(\frac{C_K^{1/\psi}h_\psi 2^{1/\psi}}{2\sqrt{\pi}}\right)\right], \tag{35}$$

$$C_K = D_K/2\kappa^2;\ D_K = \lim_{\Delta\to 0}\left[|\Delta|^{-\psi}\int\{K(x+\Delta) - K(x)\}^2dx\right], \tag{36}$$

where

$$\widehat{\mu}(\tau) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right)\Delta L_t\left(\widetilde{\theta}(\tau)\right), \tag{37}$$

$$\widetilde{\theta}(\tau) = \widehat{\theta}(\tau) - h^2\Psi\widehat{\theta}(\tau)'', \tag{38}$$

$$\widehat{\theta}(\tau) = \left[\widehat{\beta}(\tau)', \widehat{\gamma}(\tau)'\right]', \tag{39}$$

$$0 = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right)\nabla\ln f_t\left(\widehat{\beta}(\tau)\right)\ \textit{(and similarly for } \widehat{\gamma}(\tau)\textit{)}, \tag{40}$$

$\widehat{\mu}(\tau)''$ *is an estimate of the second derivative of* $\mu(\tau)$, $\widehat{\sigma}$ *is a consistent estimator of* σ *(as e.g. eq. 25 of Wu and Zhao, 2009),* $\widehat{\theta}(\tau)''$ *is an estimate of the second derivative of* $\theta(\tau)$, $1 \le \psi \le 2$ *and* h_ψ *is as in Theorem A1 of Bickel and Rosenblatt (1973) (e.g.,* $\psi = 1$ *and* $h_\psi = 1$ *for the rectangular kernel and* $\psi = 2$ *and* $h_\psi = \pi^{-1/2}$ *for the triangle, quartic, Epanechnikov and Parzen kernels).*

Corollary 5 *For the rectangular kernel, let* $m = Th$ *be an even integer. The estimator of the local relative KLIC becomes*

$$\widehat{\mu}(\tau) = \frac{1}{m}\sum_{j=[\tau T]-m/2+1}^{[\tau T]+m/2}\Delta L_j(\widetilde{\theta}(\tau)), \tag{41}$$

$[\tau T] = m/2, ..., T - m/2$, *where* $\widetilde{\theta}(\tau)$ *is the bias-adjusted local maximum likelihood estimator (38) for* $\widehat{\theta}(\tau) = \left[\widehat{\beta}(\tau)', \widehat{\gamma}(\tau)'\right]'$ *defined by*

$$0 = \frac{1}{m}\sum_{j=[\tau T]-m/2+1}^{[\tau T]+m/2}\nabla\ln f_j\left(\widehat{\beta}(\tau)\right), \tag{42}$$

(and similarly for $\widehat{\gamma}(\tau)$). The asymptotic $100(1-\alpha)\%$ simultaneous confidence bands for μ are given by

$$\widetilde{\mu}_h(\tau) \pm \widehat{\sigma}\frac{\sqrt{1/2}}{\sqrt{m}}\left[\sqrt{2\ln\left(\frac{1}{h}\right)} + \frac{1}{\sqrt{2\ln\left(\frac{1}{h}\right)}}\left[\frac{\ln\left(\ln\left(\frac{1}{h}\right)\right)}{2} + \ln\frac{1}{2\sqrt{\pi}}\right] - \frac{\ln\left[\ln(1-\alpha)^{-1/2}\right]}{\sqrt{2\ln\left(\frac{1}{h}\right)}}\right],$$

where $\widetilde{\mu}_h(\tau)$ is a bias-corrected version of $\widehat{\mu}(\tau)$ and $\widehat{\sigma}$ is a consistent estimator of σ. For example, Wu and Zhao (2007) suggest a jackknife-type bias correction scheme where $\widetilde{\mu}_h(\tau) = 2\widehat{\mu}(\tau) - \widehat{\mu}_{\sqrt{2}h}(\tau)$ and $\widehat{\mu}_{\sqrt{2}h}(\tau)$ is the estimator (41) using the bandwidth $\sqrt{2}h = \sqrt{2}m/T$ (and similarly for the parameters β, γ; e.g. $\widetilde{\beta}(\tau) = 2\widehat{\beta}(\tau) - \widehat{\beta}_{\sqrt{2}h}(\tau)$) and the long-run variance $\widehat{\sigma}$ can be estimated as

$$\widehat{\sigma} = n^{1/6}\left(2n^{2/3}-2\right)^{-1/2}\left(\sum_{i=1}^{n^{2/3}-1}\left|n^{-1/3}\sum_{j=1}^{n^{1/3}}\Delta L_{j+in^{1/3}}(\widetilde{\theta}(\tau)) - n^{-1/3}\sum_{j=1}^{n^{1/3}}\Delta L_{j+(i-1)n^{1/3}}(\widetilde{\theta}(\tau))\right|^2\right)^{1/2}. \tag{43}$$

Note that we need to correct the parameter estimate for the small sample bias typical in nonparametric estimation, and we do so by following Wu and Zhao (2009) in eq. (38). A test of the hypothesis that the models have equal performance at each point in time can be obtained by rejecting the null if the horizontal axis is not fully contained within the confidence bands obtained above.

3.3 The Fluctuation Test

In this sub-section, we consider a different measure of relative performance, which will correspond to different null and alternative hypotheses than in the previous sections. The test is based on the same nonparametric estimator of the local relative performance (31), but the difference is that we now consider an alternative asymptotic approximation in which the bandwidth is fixed instead of shrinking as the sample grows. When the bandwidth is fixed, consistent estimation of the local relative $KLIC_t$ in (15) is not possible, but what can be consistently estimated is a smoothed version of $KLIC_t$, which we call the smoothed local relative KLIC:

$$\Delta KLIC_t^*(\theta_t^*) = m^{-1}\sum_{j=t-m/2+1}^{t+m/2} E[\Delta L_j(\theta_t^*)], \quad t = m/2, ..., T-m/2, \tag{44}$$

where $\theta_t^* = (\beta_t^{*\prime}, \gamma_t^{*\prime})'$, $\Delta L_t(\theta_t^*) = \ln f_t(\beta_t^*) - \ln g_t(\gamma_t^*)$ and, e.g.,

$$\beta_t^* = \arg\max_\beta m^{-1}\sum_{j=t-m/2+1}^{t+m/2} E[f_j(\beta)], \tag{45}$$

and where $m = Th$.[8]

The null and alternative hypotheses of interest of our proposed test are:

$$\begin{aligned} H_0^{FB} &: \{\Delta KLIC_t^* = 0\} \cap \{\theta_t^* = \theta\} \text{ for } t = m/2, ..., T - m/2 \\ H_1^{FB} &: \Delta KLIC_t^* \neq 0 \text{ at some } m/2 \leq t \leq T - m/2. \end{aligned} \tag{46}$$

and a test of equal relative performance over time, which we call the fluctuation test, can be derived under the following assumptions:[9]

Assumption FB: *Let τ be s.t. $t = [\tau T]$ and $\tau \in [0, 1]$. (1) $\sigma^{-1}\left\{T^{-1/2}\sum_{j=1}^{[\tau T]}\left[\Delta L_j(\theta) - E(\Delta L_j(\theta))\right]\right\}$ obeys a Functional Central Limit Theorem (FCLT) for all $\theta \in \Theta$, Θ compact, where σ is defined in (4) below; (2) for $\widehat{\theta}_t = \left(\widehat{\beta}_t', \widehat{\gamma}_t'\right)'$, $\widehat{\beta}_t = \arg\max_\beta m^{-1}\sum_{j=t-m/2+1}^{t+m/2} f_j(\beta)$, $\widehat{\gamma}_t = \arg\max_\gamma m^{-1}\sum_{j=t-m/2+1}^{t+m/2} g_j(\gamma)$, there exists a finite and positive definite matrix V_t such that $V_t^{-1/2}\sqrt{m}\left(\widehat{\theta}_t - \theta_t^*\right) \xrightarrow{d} N(0, I)$, as $m \to \infty$ uniformly in $t > 0$, and θ_t^* is interior to Θ for every t; (3) $m^{-1}\sum_{j=t-m/2+1}^{t+m/2} \nabla \ln f_j(\beta)$ satisfy a Uniform Law of Large Numbers for $m \to \infty$ for all t, where $\nabla \ln f_j(\beta)$ is a row vector (and similarly for $\nabla \ln g_j(\gamma)$); (4) under H_0* in (46), $\sigma^2 = lim_{m\to\infty} E(m^{-1/2}\sum_{j=t-m/2+1}^{t+m/2} \Delta L_j(\theta_t^*))^2 > 0$; *(5) $m/T = h$, with $h \in (0, \infty)$ and $T \to \infty$.*

Assumption FB(4) imposes global covariance stationarity for the sequence of local likelihood differences under the null hypothesis, and it thus limits the amount of heterogeneity permitted under the null hypothesis. This assumption is in principle stronger than necessary, but it facilitates the statement of the FCLT (see Wooldridge and White, 1988 for a general FCLT for heterogeneous mixing sequences). Note that global covariance stationarity allows the variance to change over time, but in a way that ensures that, as the sample size grows, the sequence of variances converges to a finite and positive limit. For the case considered in Section 2, Assumption FB is satisfied if, for example, the parameters β_t, γ_t and the regressor variances are constant under the null hypothesis, implying that σ^2 is also constant under the null hypothesis.[10]

One can verify that Assumption FB is satisfied in the example of Section 2, where $\varepsilon_t = \Delta L_t - E[\Delta L_t] = \frac{1}{2}[\beta_t^2\left(x_t^2 - 1\right) - \gamma_t^2\left(z_t^2 - 1\right)] + (\beta_t x_t - \gamma_t z_t)\, u_t$, with variance $\sigma_t^2 = \frac{1}{2}[\beta_t^4 + \gamma_t^4] + \beta_t^2 + \gamma_t^2$. Under the null hypothesis, $\beta_t = \beta$ and $\gamma_t = \gamma$ and thus ε_t is i.i.d. and σ_t^2 is constant, which satisfies the assumptions of Donsker's FCLT theorem.

The following proposition provides a justification for the fluctuation test that we propose.

[8] For simplicity we focus here on the case of a rectangular kernel, but the definition of $\Delta KLIC_t^*$ can be extended to a general kernel.

[9] See Brown et al. (1975) and Ploberger and Kramer (1992) for fluctuation tests in the context of parameter instability.

[10] As pointed out by a referee, Assumption FB(4) can be weakened by assuming that $lim_{t\to\infty} E\sum_{j=T-m/2+1}^{t+m/2}\left[\Delta Lj(\theta_t^*)\right]^2 = \sigma_t^2 < \infty$, i.e by allowing σ_t^2 to vary across t. In this case, an estimator of σ_t^2 can be obtained as in (48) with $\widehat{\theta}_t$ instead of $\widehat{\theta}_T$, $q(t)$ instead of $q(T)$, and j ranging from $t - m/2 + q(t)$ to $t + m/2 + 1 - q(t)$, respectively.

Theorem 6 (Fluctuation Test) *Suppose Assumption FB holds. Consider the test statistic*

$$\sup_{t=m/2,...,T-m/2} F_t = \sup_{t=m/2,...,T-m/2} \widehat{\sigma}^{-1} m^{-1/2} \sum_{j=t-m/2+1}^{t+m/2} \Delta L_j(\widehat{\theta}_t), \tag{47}$$

where $\widehat{\sigma}^2$ is a HAC estimator of σ^2, given by, e.g.,

$$\widehat{\sigma}^2 = \sum_{i=-q(T)+1}^{q(T)-1} (1 - |i/q(T)|) T^{-1} \sum_{j=q(T)}^{T+1-q(T)} \Delta L_j \left(\widehat{\theta}_T\right) \Delta L_{j-i} \left(\widehat{\theta}_T\right), \tag{48}$$

$q(T)$ is a bandwidth that grows with T (e.g., Newey and West, 1987) and $\widehat{\theta}_T$ is the maximum likelihood estimator computed over the full sample. Under the null hypothesis (46)

$$F_t \Longrightarrow \left[\mathcal{B}(\tau + h/2) - \mathcal{B}(\tau - h/2)\right] / \sqrt{h}, \tag{49}$$

where $t = [\tau T]$ and $\mathcal{B}(\cdot)$ is a standard univariate Brownian motion. The critical values for a significance level α are $\pm\ k_\alpha$, where k_α solves

$$\Pr\left\{\sup_{\tau} \left|\left[\mathcal{B}(\tau + h/2) - \mathcal{B}(\tau - h/2)\right] / \sqrt{h}\right| > k_\alpha\right\} = \alpha. \tag{50}$$

The null hypothesis is rejected when $\max_t |F_t| > k_\alpha$. Simulated values of (α, k_α) are reported in Table 1 for various choices of h.

INSERT TABLE 1 HERE

4 A Small Monte Carlo Analysis

This section investigates the finite-sample size and power properties of the tests for equal performance introduced in the previous section. We consider two designs for the Data Generating Processes (DGPs), which are representative of the features discussed in the main example in Section 2. In particular, as mentioned before, the time variation in the relative KLIC might be due to the fact that the parameters change in ways that affect the local relative KLIC differently over time; design 1 focuses on this situation. However, time variation in the relative KLIC might also occur when the parameters are constant but some other aspects of the distribution of the data change in different ways over time, which will be described by design 2.

More in details, the true DGP is:

$$y_t = \beta_t x_t + \gamma_t z_t + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0,1),$$

where $x_t \sim N\left(0, \sigma^2_{x,t}\right)$, $z_t \sim N\left(0, \sigma^2_{z,t}\right)$, $t = 1, 2, ..., T$, $T = 200$. The two competing models are: Model 1: $y_t = \beta_t x_t + \varepsilon_{1,t}$ and Model 2: $y_t = \gamma_t z_t + \varepsilon_{2,t}$. We consider the following designs:

Design 1. $\sigma^2_{x,t} = \sigma^2_{z,t} = 1$, $\gamma_t = 1$, $\beta_t = 1 + \beta_A \cdot 1\,(t \leq 0.5T) - \beta_A \cdot 1\,(t > 0.5T)$. In this design, we let the parameter β change over time, and this affects the relative performance of the models over time.

Design 2. $\sigma^2_{x,t} = 1+\sigma^2_A \cdot 1\,(t > 0.75T)$, $\sigma^2_{z,t} = 1$, $\beta_t = 1$, $\gamma_t = 1$. In this design, the parameters in the conditional mean are constant but one of the variances ($\sigma^2_{x,t}$) changes over time, thus resulting in a change in the relative performance over time.

Tables 2 and 3 show the empirical rejection frequencies of the various tests for a nominal size of 5%. For the nonparametric test, we utilize a Gaussian kernel with a bandwidth equal to 0.005, which performs very well in design 1 relative to other bandwidths. Size properties are obtained by setting $\beta_A = 0$ and $\sigma_A = 0$. Table 2 demonstrates that all tests have good size properties. It also shows that the tests with highest power against a one-time reversal are the $ExpW^*_{\infty,T}$ and QLR^*_T tests; the $MeanW^*_T$ test has slightly lower power than the former. The fluctuation test has worse power properties relative to them, and the nonparametric test has considerably less power relative to all the other tests. Note that a standard full-sample likelihood ratio test would have power equal to size in design 1. Regarding design 2, Table 3 shows that, again, the nonparametric test has considerably less power than the other tests. The $ExpW^*_{\infty,T}$ and QLR^*_T tests have quite similar performance in terms of power, although the Sup-type test has slightly better power properties than the other tests, and the fluctuation test has slightly worse power properties.

INSERT TABLES 2, AND 3 HERE

Finally, Table 4 explore the robustness of our results for the nonparametric test for different bandwidth. The Monte Carlo design is the same as design 1 above. We consider a variety of bandwidths, ranging from very small ($h = 0.0005$) to quite large ($h = 0.7$). Note that the power properties do change significantly depending on the bandwidth, and that the bandwidth that performs the best is $h = 0.005$.[11]

INSERT TABLE 4 HERE

5 Empirical Application: Time-variation in the Performance of DSGE vs. BVAR Models

In a highly influential paper, Smets and Wouters (2003) (henceforth SW) show that a DSGE model of the European economy - estimated using Bayesian techniques over the period 1970:2-1999:4 - fits the data as well as atheoretical Bayesian VARs (BVARs). Furthermore, they find

[11] Unreported Monte Carlo simulations show that, however, a bandwidth that works well in one design does not necessarily work well for other designs. For example, h=0.005 is not the best choice for design 3. However, we decided to keep the bandwidth fixed across Monte Carlo designs, as the researcher does not know the DGP in practice.

that the parameter estimates from the DSGE model have the expected sign. Perhaps for these reasons, this new generation of DSGE models has attracted a lot of interest from forecasters and central banks. SW's model features include sticky prices and wages, habit formation, adjustment costs in capital accumulation and variable capacity utilization, and the model is estimated using seven variables: GDP, consumption, investment, prices, real wages, employment, and the nominal interest rate. Their conclusion that the DSGE fits the data as well as BVARs is based on the fact that the marginal data densities for the two models are of comparable magnitudes over the full sample. However, given the changes that have characterized the European economy over the sample analyzed by SW - for example, the creation of the European Union in 1993, changes in productivity and in the labor market, to name a few - it is plausible that the relative performance of theoretical and atheoretical models may itself have varied over time. In this section, we apply the techniques proposed in this paper to assess whether the relative performance of the DSGE model and of BVARs was stable over time. We extend the sample considered by SW to include data up to 2004:4, for a total sample of size $T = 145$.

In order to compute the local measure of relative performance, we estimate both models recursively over a moving window of size $m = 70$ using Bayesian methods. As in SW, the first 40 data points in each sample are used to initialize the estimates of the DSGE model and as training samples for the BVAR priors. We consider a BVAR(1) and a BVAR(2), both of which use a variant of the Minnesota prior, as suggested by Sims (2003).[12] We present results for two different transformations of the data. The first applies the same detrending of the data used by SW, which is based on a linear trend fitted on the whole sample (we refer to this as "full-sample detrending"). As cautioned by Sims (2003), this type of pre-processing of the data may unduly favour the DSGE, and thus we further consider a second transformation of the data, where detrending is performed on each rolling estimation window ("rolling-sample detrending").

Figure 2 displays the evolution of the posterior mode of some representative parameters. Figure 2(a) shows parameters that describe the evolution of the persistence of some representative shocks (productivity, investment, government spending, and labor supply); Figure 2(b) shows the estimates of the standard deviation of the same shocks; and Figure 2(c) plots monetary policy parameters. Overall, Figure 2 reveals evidence of parameter variation. In particular, the figures show some decrease in the persistence of the productivity shock, whereas both the persistence and the standard deviation of the investment shock seem to increase over time. The monetary policy parameters appear to be overall stable over time.

FIGURE 2 HERE

[12] The BVAR's were estimated using software provided by Chris Sims at www.princeton.edu/~sims. As in Sims (2003), for the Minnesota prior we set the decay parameter to 1 and the overall tightness to .3. We also included sum-of-coefficients (with weight $\mu = 1$) and co-persistence (with weight $\lambda = 5$) prior components.

We then apply our fluctuation test to test the hypothesis that the DSGE model and the BVAR have equal performance at every point in time over the historical sample.

Figure 3 shows the implementation of the fluctuation test for the DSGE vs. a BVAR(1) and BVAR(2), using full-sample detrending of the data. The estimate of the local relative $KLIC$ is evaluated at the posterior modes $\widehat{\beta}_t$ and $\widehat{\gamma}_t$ of the models' parameters, using the fact that $\widehat{\beta}_t$ and $\widehat{\gamma}_t$ are consistent estimates of the pseudo-true parameters β_t^* and γ_t^* (see, e.g., Fernandez-Villaverde and Rubio-Ramirez, 2004).

FIGURE 3 HERE

Figure 3 suggests that the DSGE has comparable performance to both a BVAR(1) and BVAR(2) up until the early 1990s, at which point the performance of the DSGE dramatically improves relative to that of the reduced-form models.

To assess whether this result is sensitive to the data filtering, we implement the fluctuation test for the DSGE vs. a BVAR(1) and BVAR(2), this time using rolling-window detrended data.

FIGURE 4 HERE

The results confirm the suspicion expressed by Sims (2003) that the pre-processing of the data utilized by SW penalizes the reduced-form models in favour of the DSGE. As we see from Figure 4, once the detrending is performed on each rolling window, the advantage of the DSGE at the end of the sample disappears, and the DSGE performs as well as a BVAR(1) on most of the sample, whereas it is outperformed by a BVAR(2) for all but the last few dates in the sample (when the two models perform equally well).

6 Conclusions

This paper developed statistical testing procedures for evaluating models' relative performance in unstable environments. We proposed three tests: 1) a one-time reversal test; 2) a nonparametric test; and 3) a fluctuation test. We investigated the advantages and limitations of the different approaches and compared the quality of the approximation that they deliver in finite samples. Based on the results of the latter, we do not recommend the nonparametric test for typical macroeconomic applications, whereas the choice between the one-time reversal and the fluctuation test should be driven by the type of alternative hypothesis of interest in a given application. Finally, an empirical application to the European economy points to the presence of instabilities in the models' parameters, and suggests that a VAR fitted the last two decades of data better than a standard DSGE model, a conclusion that is however sensitive to the detrending method utilized.

References

[1] Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica* 59, 817-858.

[2] Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica* 61, 821–856.

[3] Andrews, D.W.K., and W. Ploberger (1994), "Optimal Tests When a Nuisance Parameter is Present only under the Alternative", *Econometrica* 62(6), 1383-1414.

[4] Bickel, P.J. and M. Rosenblatt (1973), "On Some Global Measures of the Deviations of Density Function Estimates", *Annals of Statistics* 1, 1071-1095.

[5] Brown, R.L., J. Durbin and J.M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships over Time with Comments", *Journal of the Royal Statistical Society, Series B*, 37, 149-192.

[6] Cavaliere, G. and R. Taylor (2005), "Stationarity Tests Under Time-Varying Second Moments", *Econometric Theory* 21, 1112-1129.

[7] Domowitz, I. and H. White (1982), "Mis-specified Models with Dependent Observations," *Journal of Econometrics* 20(1), 35-58.

[8] Elliott, G. and U. Muller (2006), "Efficient Tests for General Persistent Time Variation in Regression Coefficients", *The Review of Economic Studies* 73*,* 907-940.

[9] Fernández-Villaverde, J. and J. Rubio-Ramírez (2004), "Comparing Dynamic Equilibrium. Models to Data: A Bayesian Approach.", *Journal of Econometrics* 123.

[10] Giacomini, R. and B. Rossi (2010), "Forecast Comparisons in Unstable Environments", *Journal of Applied Econometrics,* in press.

[11] Muller, U., "Efficient Tests under a Weak Convergence Assumption", *mimeo*, Princeton University.

[12] Muller, U. and P. Petalas (2009), "Efficient Estimation of the Parameter Path in Unstable Time Series Models", *The Review of Economic Studies*, forthcoming.

[13] Newey, W., and K. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708.

[14] Ploberger, W. and W. Kramer (1992), "The Cusum Test with Ols Residuals", *Econometrica* 60(2), 271-285.

[15] Qu, Z. and P. Perron (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica* 75(2), 459-502.

[16] Rivers, D. and Q. Vuong (2002), "Model Selection Tests for Nonlinear Dynamic Models", *Econometrics Journal*, 5, 1-39.

[17] Rossi, B. (2005), "Optimal Tests for Nested Model Selection with Underlying Parameter Instabilities", *Econometric Theory* 21(5), 962-990.

[18] Sims, C. (2003), "Comment on Smets and Wouters", *mimeo*, available at: https://sims.princeton.edu/yftp/Ottawa/SWcommentSlides.pdf

[19] Smets, F. and R. Wouters (2003), "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area", *Journal of the European Economic Association,* 1, 1123-1175.

[20] Stock, J.H., and M.W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices", *Journal of Economic Literature.*

[21] Van der Vaart, A. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer-Verlag: New York.

[22] Vuong, Q. H. (1989), "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses", *Econometrica*, 57, 307-333.

[23] Wooldridge, J.M. and H. White (1988), "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory* 4, 210-230.

[24] Wu, W.B., and Z. Zhao (2007), "Inference of Trends in Time Series", *Journal of the Royal Statistical Society* B69, 391-410.

7 Appendix A - Proofs

Lemma 7 *Let the approximate ML estimators be*

$$\overline{\mu}_1(\delta) \equiv T^{-1/2} \sum_{t=1}^{[T\delta]} \Delta L_t(\beta_1, \gamma_1) \tag{51}$$

and

$$\overline{\mu}_2(\delta) \equiv T^{-1/2} \sum_{t=[T\delta]+1}^{T} \Delta L_t(\beta_2, \gamma_2). \tag{52}$$

Under Assumption OT, $sup_{\delta\in\Pi}\left[\sqrt{T}\widehat{\mu}_1(\delta) - \overline{\mu}_1(\delta)\right] \underset{p}{\rightarrow} 0$ *and* $sup_{\delta\in\Pi}[\sqrt{T}\widehat{\mu}_2(\delta) - \overline{\mu}_2(\delta)] \underset{p}{\rightarrow} 0$ *under* H_0.

Proof of Lemma (7). For every $\delta \in \Pi$, let:

$$\begin{aligned}
\widehat{\mu}_1(\delta) &= \frac{1}{[T\delta]} \sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\beta}_1(\delta), \widehat{\gamma}_1(\delta)\right) &(53)\\
\widehat{\mu}_2(\delta) &= \frac{1}{T - [\delta T]} \sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\beta}_2(\delta), \widehat{\gamma}_2(\delta)\right). &(54)
\end{aligned}$$

From a mean value expansion of (53):

$$\begin{aligned}
\sqrt{T}\widehat{\mu}_1(\delta) - \overline{\mu}_1(\delta) &= \sqrt{T}\left[T^{-1}\sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\beta}_1(\delta), \widehat{\gamma}_1(\delta)\right) - T^{-1}\sum_{t=1}^{[T\delta]} \Delta L_t(\beta_1, \gamma_1)\right] \\
&= T^{-1/2}\sum_{t=1}^{[T\delta]} E\left(\frac{\partial}{\partial\beta_1(\delta)} \Delta L_t(\theta_1)\right)'\left(\widehat{\beta}_1(\delta) - \beta_1\right) \\
&\quad + T^{-1/2}\sum_{t=1}^{[T\delta]} E\left(\frac{\partial}{\partial\gamma_1(\delta)} \Delta L_t(\theta_1)\right)'\left(\widehat{\gamma}_1(\delta) - \gamma_1\right)
\end{aligned}$$

$$\begin{aligned}
&+T^{-1/2}\sum_{t=1}^{[T\delta]}\left[\frac{\partial}{\partial\beta_1(\delta)}\Delta L_t\left(\ddot{\beta}_1(\delta), \gamma_1\right) - E\left(\frac{\partial}{\partial\beta_1(\delta)}\Delta L_t(\beta_1, \gamma_1)\right)\right]'\left(\widehat{\beta}_1(\delta) - \beta_1\right) \\
&+T^{-1/2}\sum_{t=1}^{[T\delta]}\left[\frac{\partial}{\partial\gamma_1(\delta)}\Delta L_t\left(\beta_1, \ddot{\gamma}_1(\delta)\right) - E\left(\frac{\partial}{\partial\gamma_1(\delta)}\Delta L_t(\beta_1, \gamma_1)\right)\right]'\left(\widehat{\gamma}_1(\delta) - \gamma_1\right),
\end{aligned}$$

where $\ddot{\beta}_1(\delta)$ is an intermediate point between $\widehat{\beta}_1(\delta)$ and β_1, and similarly for $\ddot{\gamma}_1(\delta)$. The last two terms are $o_p(1)$ by Assumptions OT(3) and OT(4). The first two terms in the equality are $o_p(1)$ because $\left(\widehat{\beta}_1(\delta) - \beta_1\right)T^{1/2}$ is $O_p(1)$ by Assumption OT(4) and $E\left(\frac{\partial}{\partial\beta_1(\delta)}\Delta L_t(\theta_1)\right) = 0$ (and similarly for $\frac{\partial}{\partial\gamma_1}\Delta L_t(.)$). The result follows similarly for (54). ■

Proof of Theorem 1. Define the approximate $ExpW_T$ for testing $H_0^{OT}: E[\Delta L_t(\theta)] = 0$ as:

$$\begin{aligned} Exp\overline{W}_{c,T} &= (1+c)^{-1/2}\int \exp\left(\frac{1}{2}\frac{c}{1+c}\overline{W}_{c,T}(\delta)\right) dJ(\delta), \\ \text{where } \overline{W}_{c,T}(\delta) &= \overline{\mu}(\delta)' H' \left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1} H\overline{\mu}(\delta). \end{aligned}$$

and define the approximate LR statistic for testing H_0^{OT} as:

$$\begin{aligned} \overline{LR}_T &= \int \exp\left(\frac{1}{2}\left[H\overline{\mu}(\delta)\right]' \left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}\left[H\overline{\mu}(\delta)\right]\right)\cdot \\ &\int \exp\left(-\frac{1}{2}\left[H\left(\overline{\mu}(\delta)-\mu^*\right)\right]' \left(H\mathcal{I}_{0,\delta}^{-1}H\right)^{-1}\left[H\left(\overline{\mu}(\delta)-\mu^*\right)\right]\right) dQ_\delta(\mu^*)\, dJ(\delta). \end{aligned} \tag{55}$$

To prove Part (i) in the Theorem, we will show: (a) $LR_T - \overline{LR}_T \underset{p}{\to} 0$, (b) $\overline{LR}_T = Exp\overline{W}_{c,T}$, (c) $Exp\overline{W}_{c,T} - ExpW_T^* \underset{p}{\to} 0$.

(a) Follows from Andrews and Ploberger's (1994) Lemma A2 under Assumptions OT(2), OT(3), OT(4) and OT(5).

(b) Let $\widehat{\mu}(\delta) \equiv \left[\widehat{\mu}_1(\delta)',\ \widehat{\mu}_2(\delta)'\right]'$, where

$$\widehat{\mu}_1(\delta) = \frac{1}{[T\delta]}\sum_{t=1}^{[T\delta]}\Delta L_t\left(\widehat{\theta}_1(\delta)\right),\ \widehat{\mu}_2(\delta) = \frac{1}{[T(1-\delta)]}\sum_{t=[T\delta]+1}^{T}\Delta L_t\left(\widehat{\theta}_2(\delta)\right). \tag{56}$$

Let approximate estimators be defined as (51) and (52). Lemma 7 shows that the approximate estimators $\overline{\mu}_1(\delta), \overline{\mu}_2(\delta)$ are asymptotically equivalent to $\widehat{\mu}_1(\delta), \widehat{\mu}_2(\delta)$.

From (55) and Assumption OT(5), which implies $Q_\delta(H\mu^*) = N\left(0, cH\mathcal{I}_{0,\delta}^{-1}H'\right)$, we have:

$$\begin{aligned} \overline{LR}_T &= \int \exp\left(\frac{1}{2}\left[H\overline{\mu}(\delta)\right]' \left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}\left[H\overline{\mu}(\delta)\right]\right)\cdot \\ &\int \exp\left(-\frac{1}{2}\left[H\left(\overline{\mu}(\delta)-\mu^*\right)\right]' \left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}\left[H\left(\overline{\mu}(\delta)-\mu^*\right)\right]\right)\frac{1}{\sqrt{2\delta}}\det\left(c^{-1/2}\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1/2}\right) \\ &\times\exp\left(-\frac{1}{2}\mu^{*\prime}H'\left(cH\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\mu^*\right) d\mu^* dJ(\delta) \\ &= \int\int \exp\left(\frac{1}{2}\left(1+\frac{1}{c}\right)\left[-\mu^{*\prime}H'\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\mu^* + 2\left(1+\frac{1}{c}\right)^{-1}\overline{\mu}(\delta)'H'\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\mu^*\right]\right) \\ &\times\frac{1}{\sqrt{2\delta}}\det\left(c^{-1/2}\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1/2}\right) d\mu^* dJ(\delta) \end{aligned}$$

$$\begin{aligned} &= \int\int \exp\left(-\frac{1}{2}\left(1+\frac{1}{c}\right)\left\{\left[\mu^* - \left(1+\frac{1}{c}\right)^{-1}\overline{\mu}(\delta)\right]' H'\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\left[\mu^* - \left(1+\frac{1}{c}\right)^{-1}\overline{\mu}(\delta)\right]\right.\right. \\ &\left.\left. -\left(1+\frac{1}{c}\right)^{-2}\overline{\mu}(\delta)'H\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H'\overline{\mu}(\delta)\right\}\right) d\mu^*\frac{1}{\sqrt{2\delta}}\det\left(c^{-1/2}\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1/2}\right) dJ(\delta) \\ &= \int\left[\int \frac{1}{\sqrt{2\delta\left(1+\frac{1}{c}\right)^{-1}}}\det\left(\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1/2}\right)\exp\left\{-\frac{1}{2}\left(1+\frac{1}{c}\right)\left[\mu^* - \left(1+\frac{1}{c}\right)^{-1}\overline{\mu}(\delta)\right]'\right.\right. \end{aligned}$$

$$
\begin{aligned}
& \times H\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\left[\mu^{*}-\left(1+\frac{1}{c}\right)^{-1}\overline{\mu}\left(\delta\right)\right]\right\}d\mu^{*}]\Bigg] \\
& \times\exp\left\{\frac{1}{2}\left(1+\frac{1}{c}\right)\left(1+\frac{1}{c}\right)^{-2}\overline{\mu}\left(\delta\right)'H'\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\overline{\mu}\left(\delta\right)\right\}\sqrt{\left(1+\frac{1}{c}\right)^{-1}}\frac{1}{\sqrt{c}}dJ\left(\delta\right) \\
= & \int\exp\left\{\frac{1}{2}\left(1+\frac{1}{c}\right)^{-1}\overline{\mu}\left(\delta\right)'H'\left(H\mathcal{I}_{0,\delta}^{-1}H'\right)^{-1}H\overline{\mu}\left(\delta\right)\right\}\frac{1}{\sqrt{1+c}}dJ\left(\delta\right)
\end{aligned}
$$

(c) From Lemma 7 we have $sup_{\delta\in\Pi}\left[\sqrt{T}\widehat{\mu}_1\left(\delta\right)-\overline{\mu}_1\left(\delta\right)\right]\underset{p}{\rightarrow}0$. Also, note that $2\left[\xi\left(\widehat{\theta}\left(\delta\right),\delta\right)-\xi\left(0\right)\right]=\left(\widehat{\theta}\left(\delta\right)-\theta\right)'\frac{\partial^2\xi(0)}{\partial\theta\partial\theta'}\left(\widehat{\theta}\left(\delta\right)-\theta\right)+o_p\left(1\right)=\mu\left(\widehat{\theta}\left(\delta\right)\right)'M^{-1\prime}\frac{\partial^2\xi(0)}{\partial\theta\partial\theta'}M^{-1}\mu\left(\widehat{\theta}\left(\delta\right)\right)+o_p\left(1\right),$ where $M=\frac{\partial\mu(\theta)}{\partial\theta\partial\theta'}$.

Part (ii) follows from Lemma A4 in Andrews and Ploberger (1994), which guarantees that for the local alternatives in (20), the density $l_t\left(\mu^{*}T^{-1/2},\delta\right)$ is contiguous to the density $l_t\left(0,\delta\right)$. ■

Proof of Corollary 2. Follows from Theorem 1, where (28), (29) follow from Assumption OT(1). ■

Proof of Proposition 3. First we show that: (i) $LM_1=\sigma^{-2}T^{-1/2}\left[\sum_{t=1}^{T}\left(\ln f_t(\beta_t)-\ln g_t\left(\gamma_t\right)\right)\right]^2+o_p\left(1\right)$ and (ii)

$$
\begin{aligned}
LM_2\left(\delta\right) = \; & \sigma^{-2}\left(\left[\delta T\right]/T\right)^{-1}\left(1-\left[\delta T\right]/T\right)^{-1} \\
& [T^{-1/2}\sum_{t=1}^{[\delta T]}\left(\ln f_t(\beta_t)-\ln g_t\left(\gamma_t\right)\right)+ \\
& -\left(\left[\delta T\right]/T\right)T^{-1/2}\sum_{t=1}^{T}\left(\ln f_t(\beta_t)-\ln g_t\left(\gamma_t\right)\right)]^2+o_p\left(1\right)
\end{aligned}
$$

To prove (i), consider

$$
LM_1=\sigma^{-2}T^{-1/2}\sum_{t=1}^{T}\left(\ln f_t(\widehat{\beta}_T\left(\delta\right))-\ln g_t\left(\widehat{\gamma}_T\left(\delta\right)\right)\right) \tag{57}
$$

Consider the first summand in (57). By applying a Taylor expansion around $\theta = \left[\beta', \gamma'\right]'$:

$$
\begin{aligned}
& \sigma^{-2} T^{-1/2} \sum_{t=1}^{T} \left(\ln f_t(\widehat{\beta}_T(\delta)) - \ln g_t(\widehat{\gamma}_T(\delta)) \right) \\
= \; & \sigma^{-2} T^{-1/2} \sum_{t=1}^{[T\delta]} \left(\ln f_t(\beta(\delta)) - \ln g_t(\gamma(\delta)) \right) + \\
& + \frac{1}{2} \sigma^{-2} T^{-1/2} \sum_{t=1}^{[T\delta]} \left[\nabla \ln f_t(\ddot{\beta}_T(\delta)) - E\left(\nabla \ln f_t(\ddot{\beta}_T(\delta)) \right) \right] \left(\widehat{\beta}_T(\delta) - \beta(\delta) \right) \\
& - \frac{1}{2} \sigma^{-2} T^{-1/2} \sum_{t=1}^{[T\delta]} \left[\nabla \ln g_t\left(\ddot{\gamma}_T(\delta)\right) - E\left(\nabla \ln g_t\left(\ddot{\gamma}_T(\delta)\right) \right) \right] \left(\widehat{\gamma}_T(\delta) - \gamma(\delta) \right) \\
& \frac{1}{2} \sigma^{-2} T^{-1/2} \sum_{t=1}^{[T\delta]} E\left(\nabla \ln f_t(\ddot{\beta}_T(\delta)) \right) \left(\widehat{\beta}_1(\delta) - \beta(\delta) \right) \\
& - \frac{1}{2} \sigma^{-2} T^{-1/2} \sum_{t=1}^{[T\delta]} E\left(\nabla \ln g_t\left(\ddot{\gamma}_T(\delta)\right) \right) \left(\widehat{\gamma}_1(\delta) - \gamma(\delta) \right) \\
= \; & \sigma^{-2} T^{-1/2} \sum_{t=1}^{[T\delta]} \left(\ln f_t(\beta(\delta)) - \ln g_t(\gamma(\delta)) \right) + o_p(1) + o_p(1)
\end{aligned}
$$

where $\ddot{\beta}_T(\delta)$ is an intermediate point between $\widehat{\beta}_T(\delta)$ and $\beta(\delta)$ (similarly for $\ddot{\gamma}_T(\delta)$). Assumption OT(3) ensures that $T^{-1} \sum_{t=1}^{T} \left[\nabla \ln f_t(\ddot{\beta}_T(\delta)) - E\left(\nabla \ln f_t(\ddot{\beta}_T(\delta)(\delta)) \right) \right] \underset{p}{\rightarrow} 0$ under the null hypothesis (and similarly for the component in γ) and, by Assumption OT(2), $T^{-1/2}\left(\widehat{\beta}_T(\delta) - \beta(\delta) \right) = O(1)$ which proves the first $o_p(1)$. The second $o_p(1)$ is justified by the fact that $E\left(\nabla \ln f_t(\ddot{\beta}_T(\delta)) \right) - E\left(\nabla \ln f_t(\beta(\delta)) \right) \underset{p}{\rightarrow} 0$, where $E\left(\nabla \ln f_t(\beta(\delta)) \right) = 0$, and that $T^{-1/2}\left(\widehat{\beta}_T(\delta) - \beta(\delta) \right) = O(1)$. A similar argument proves that this results holds for the second summand as well as for (ii).

By Assumption OT(1), under the null hypothesis:

$$
\sigma^{-1} T^{-1/2} \sum_{t=1}^{T} \left(\ln f_t(\beta_t) - \ln g_t(\gamma_t) \right) \Longrightarrow \mathcal{B}(1) \tag{58}
$$

$$
\begin{aligned}
& \sigma^{-1} \left([\delta T] / T \right)^{-1/2} \left(1 - [\delta T] / T \right)^{-1/2} [T^{-1/2} \sum_{t=1}^{[\delta T]} \left(\ln f_t(\beta_t) - \ln g_t(\gamma_t) \right) \\
& - \left([\delta T] / T \right) T^{-1/2} \sum_{t=1}^{T} \left(\ln f_t(\beta_t) - \ln g_t(\gamma_t) \right)] \\
\Longrightarrow \quad & \delta^{-1/2} (1-\delta)^{-1/2} \left[\mathcal{B}(\delta) - \delta \mathcal{B}(1) \right] = \delta^{-1/2} (1-\delta)^{-1/2} \mathcal{BB}(\delta)
\end{aligned} \tag{59}
$$

where (58) and (59) are asymptotically independent. Then:

$$\begin{aligned} LM_1 + LM_2(\delta) &= \sigma^{-2} T^{-1/2} \left[\sum_{t=1}^{T} (\ln f_t(\beta_t) - \ln g_t(\gamma_t)) \right]^2 \\ &\quad + \sigma^{-2} \left(\frac{[T\delta]}{T} \right)^{-1} \left(1 - \frac{[T\delta]}{T} \right)^{-1} [T^{-1/2} \sum_{t=1}^{[T\delta]} (\ln f_t(\beta_t) - \ln g_t(\gamma_t)) \\ &\quad - \left(\frac{t}{T} \right) T^{-1/2} \sum_{t=1}^{T} (\ln f_t(\beta_t) - \ln g_t(\gamma_t))]^2 + o_p(1) \\ &\Longrightarrow \mathcal{B}(1)^2 + \delta^{-1} (1-\delta)^{-1} \mathcal{BB}(\delta)^2 \end{aligned}$$

and the result follows by the Continuous Mapping Theorem. ■

Proof of Proposition 4. In this proof, let $\widehat{\mu}(\tau)$ in (37) be denoted by $\widehat{\mu}\left(\tau, \widetilde{\theta}(\tau)\right)$ to emphasize its dependence on the estimated parameters, $\widetilde{\theta}(\tau)$. The confidence bands in (33) are obtained by showing that for every $u \in \mathbb{R}$ and for $T \to \infty$,

$$\Pr\left\{ \frac{\sqrt{Th}}{\sigma} \sup_{\tau \in [wh, 1-wh]} \left[\widehat{\mu}\left(\tau, \widetilde{\theta}(\tau)\right) - \mu(\tau, \theta(\tau)) - h^2 \Psi \mu''(\tau, \theta(\tau)) - B_K \right] \leq \frac{u}{\sqrt{2\ln(1/h)}} \right\} \to \exp(-2\exp(-u)). \tag{60}$$

We have

$$\begin{aligned} & \widehat{\mu}\left(\tau, \widetilde{\theta}(\tau)\right) - \mu(\tau, \theta(\tau)) - h^2 \Psi \mu''(\tau, \theta(\tau)) \\ = \; & \underbrace{\widehat{\mu}\left(\tau, \widetilde{\theta}(\tau)\right) - E[\widehat{\mu}(\tau, \theta(\tau))]}_{A} + \underbrace{E[\widehat{\mu}(\tau, \theta(\tau))] - \left(\mu(\tau, \theta(\tau)) - h^2 \delta \mu''(\tau, \theta(\tau))\right)}_{B}. \end{aligned}$$

From Lemma 3 of Wu and Zhao (2007), the term B is $O\left(h^3 + T^{-1}h^{-1}\right)$ uniformly over $\tau \in [\omega h, 1-\omega h]$ by Assumption SB3, which implies that the term is asymptotically negligible**,** and thus one need only to focus on term A and show that

$$\Pr\left\{ \frac{\sqrt{Th}}{\sigma} \sup_{\tau \in [wh, 1-wh]} \left[\widehat{\mu}\left(\tau, \widetilde{\theta}(\tau)\right) - E[\widehat{\mu}(\tau, \theta(\tau))] - B_K \right] \leq \frac{u}{\sqrt{2\ln(1/h)}} \right\} \to \exp(-2\exp(-u)).$$

To prove this, define

$$\begin{aligned} \widehat{U}_T(\tau) &\equiv \sqrt{Th}\sigma^{-1} \left[\widehat{\mu}\left(\tau, \widetilde{\theta}(\tau)\right) - \mu_t \right] \\ &= \sigma^{-1} \left\{ \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h} \right) \left[\Delta L_t\left(\widetilde{\theta}(\tau)\right) - \mu_t \right] \right\} \\ &= \sigma^{-1} \left\{ \frac{1}{\sqrt{Th}} \sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h} \right) \left[\Delta L_t\left(\widetilde{\theta}(\tau)\right) - \Delta L_t(\theta(\tau)) \right] \right\} + U_T(\tau) \end{aligned} \tag{61}$$

where $U_T(\tau) \equiv \frac{1}{\sqrt{Th}}\sigma^{-1}\left\{\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\varepsilon_t\right\}$. By a Mean Value expansion of $\Delta L_t\left(\widetilde{\theta}(\tau)\right)$ around $\Delta L_t(\theta(\tau))$, we have:

$$\begin{aligned}
\widehat{\mu}\left(\tau,\widetilde{\theta}(\tau)\right) &\equiv \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta L_t\left(\widetilde{\theta}(\tau)\right) = \qquad (62)\\
&= \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta L_t(\theta(\tau))\\
&\quad + \left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\\
&= \widehat{\mu}(\tau,\theta(\tau)) + \left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)
\end{aligned}$$

where $\overline{\theta}(\tau)$ lies between $\widetilde{\theta}(\tau)$ and $\theta(\tau)$.

By substituting (62) in (61), we have:

$$\widehat{U}_T(\tau) = U_T(\tau) + \left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)$$

Let $W_T(\tau) \equiv \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta B(t)$. Note that

$$\begin{aligned}
& \ln^{1/2}(T)\,|\widehat{U}_T(\tau) - W_T(\tau)|\\
\leq\ & \ln^{1/2}(T)\,|U_T(\tau) - W_T(\tau)| + \ln^{1/2}(T)\left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right|\\
=\ & o_p(1) + \ln^{1/2}(T)\left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right| = o_p(1) \qquad (63)
\end{aligned}$$

where the first inequality follows from the triangle inequality, the first $o_p(1)$ in the second equality follows from Wu and Zhao (2007, Lemma 2, eq. 29), and the last $o_p(1)$ in (63) follows from the fact that

$$\begin{aligned}
& \ln^{1/2}(T)\left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right|\\
\leq\ & \left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\right|\cdot\left|\sqrt{Th}\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right|\cdot\sqrt{\frac{\ln(T)}{Th}}
\end{aligned}$$

and $\left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\right| = O_p(1)$ by Assumption SB(5), $E[\Delta s_t(\theta(\tau))] = 0$ and consistency of $\widetilde{\theta}(\tau)$ for $\theta(\tau)$; $\left|\sqrt{Th}\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right| = O_{as}(1)$ by Assumption SB(7); and $\sqrt{\frac{\ln(T)}{Th}} =$

$o(1)$ by Assumption SB(3). In practice, the bias-corrected estimator $\widetilde{\mu}(\tau)$ can be obtained with a jackknife-type bias correction scheme where $\widetilde{\mu}(\tau) = 2\widehat{\mu}(\tau) - \widehat{\mu}_{\sqrt{2}h}(\tau)$, where, again, the estimation uncertainty on $\widetilde{\theta}(\tau)$ is irrelevant, as above. The consistency of the proposed estimator $\widehat{\sigma}$ follows from the discussion of eq. (25) of Wu and Zhao (2007) and the consistency of the estimated parameters, $\widetilde{\theta}(\tau)$, using arguments similar to the above. ■

Proof of Theorem 6. Let $\sum_j \equiv \sum_{j=t-m/2+1}^{t+m/2}$ for $t = m/2, ..., T - m/2$. We first show that $\sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t) = \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) + o_p(1)$. Applying a mean value expansion, we have:

$$
\begin{aligned}
& \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t) \qquad (64)\\
= \; & \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) \\
& +\sigma^{-1}\frac{1}{2}\left\{\left\{m^{-1}\sum_j \nabla \ln f_j(\ddot{\beta}_t) - E\left[m^{-1}\sum_j \nabla \ln f_j(\ddot{\beta}_t)\right]\right\}\sqrt{m}\left(\widehat{\beta}_t - \beta^*\right)\right. \\
& \left. -\left\{m^{-1}\sum_j \nabla \ln f_j(\ddot{\gamma}_t) - E\left[m^{-1}\sum_j \nabla \ln g_j(\ddot{\gamma}_t)\right]\right\}\sqrt{m}\left(\widehat{\gamma}_t - \gamma^*\right)\right\} \\
& +\sigma^{-1}\frac{1}{2}\left\{E\left[m^{-1}\sum_j \nabla \ln f_j(\ddot{\beta}_t)\right]\sqrt{m}\left(\widehat{\beta}_t - \beta^*\right)\right. \\
& \left. -E\left[m^{-1}\sum_j \nabla \ln g_j(\ddot{\gamma}_t)\right]\sqrt{m}\left(\widehat{\gamma}_t - \gamma^*\right)\right\} \\
= \; & \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) + o_p(1) + o_p(1),
\end{aligned}
$$

where $\ddot{\beta}_t$ is an intermediate point between $\widehat{\beta}_t$ and β^* (and similarly for $\ddot{\gamma}_t$).By Assumption FB(3) we have that $m^{-1}\sum_j \nabla \ln f_j(\ddot{\beta}_t) - E\left[m^{-1}\sum_j \nabla \ln f_j(\ddot{\beta}_t)\right]$ is $o_p(1)$ and, by Assumption FB(2) and under H_0, $\sqrt{m}\left(\widehat{\beta}_t - \beta^*\right)$ is $O_p(1)$ (and similarly for the second model), which proves the first $o_p(1)$ in equation (64). The second $o_p(1)$ follows from the fact that $\sqrt{m}\left(\widehat{\beta}_t - \beta^*\right)$ is $O_p(1)$ by FB(2) and that $E\left[m^{-1}\sum_j \nabla \ln f_j(\ddot{\beta}_t)\right] - E\left[m^{-1}\sum_j \nabla \ln f_j(\beta^*)\right] \underset{p}{\rightarrow} 0$ as $m \rightarrow \infty$ by Theorem 2.3 of Domowitz and White (1982) given Assumptions FB(2) and FB(3), where $E\left[m^{-1}\sum_j \nabla \ln f_j(\beta^*)\right] = 0$ by definition. Now write

$$
\begin{aligned}
& \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) \\
= \; & (m/T)^{-1/2}\left(\sigma^{-1}T^{-1/2}\sum_{j=1}^{t+m/2}\Delta L_j(\theta^*) - \sigma^{-1}T^{-1/2}\sum_{j=1}^{t-m/2}\Delta L_j(\theta^*)\right).
\end{aligned}
$$

By Assumptions FB(1), FB(4) and FB(5), we have

$$\sigma^{-1}m^{-1/2}\sum_{j}\Delta L_j(\theta^*) \Longrightarrow \left[\mathcal{B}\left(\tau+h/2\right)-\mathcal{B}\left(\tau-h/2\right)\right]/\sqrt{h},$$

where $t=[\tau T]$. The statement in the proposition then follows from the fact that, under H_0, $\widehat{\sigma}$ in (48) is a consistent estimator of σ. ■

8 Tables and Figures

Table 1. Critical values for the fluctuation test (k_α)

	α	
h	0.05	0.10
0.1	3.393	3.170
0.2	3.179	2.948
0.3	3.012	2.766
0.4	2.890	2.626
0.5	2.779	2.500
0.6	2.634	2.356
0.7	2.560	2.252
0.8	2.433	2.130
0.9	2.248	1.950

Notes to Table 1. The table reports critical values for the fluctuation test in Proposition 6. Values of k_α in Table 1 are obtained by Monte Carlo simulations (based on 8,000 Monte Carlo replications and by approximating the Brownian motion with 400 observations)

Table 2. Monte Carlo: Design 1

β_A	Nonparametric	Fluctuation	QLR^*_T	Break	$\text{ExpW}^*_{\infty,T}$	MeanW^*_T
0	0.06	0.04	0.04	0.05	0.04	0.05
0.1	0.05	0.06	0.09	0.10	0.09	0.08
0.2	0.06	0.15	0.20	0.24	0.19	0.16
0.3	0.08	0.32	0.44	0.51	0.42	0.34
0.4	0.11	0.53	0.69	0.75	0.66	0.56
0.5	014	0.72	0.86	0.90	0.84	0.76
0.6	0.19	0.87	0.96	0.97	0.95	0.90
0.7	0.27	0.94	0.99	0.99	0.99	0.96
0.8	0.34	0.98	1	1	1	0.98
0.9	0.42	0.99	1	1	1	1
1.0	0.50	1	1	1	1	1
1.1	0.58	1	1	1	1	1
1.2	0.68	1	1	1	1	1
1.3	0.74	1	1	1	1	1
1.4	0.78	1	1	1	1	1
1.5	0.86	1	1	1	1	1
1.6	0.90	1	1	1	1	1
1.7	0.94	1	1	1	1	1
1.8	0.96	1	1	1	1	1
1.9	0.98	1	1	1	1	1
2	0.99	1	1	1	1	1

Table 3. Monte Carlo: Design 2

σ_A^2	Nonparametric	Fluctuation	QLR_T^*	Break	$\text{ExpW}_{\infty,T}^*$	MeanW_T^*
0	0.05	0.04	0.05	0.05	0.05	0.05
0.1	0.05	0.05	0.06	0.06	0.06	0.06
0.2	0.06	0.06	0.10	0.10	0.10	0.09
0.3	0.08	0.08	0.18	0.16	0.16	0.14
0.4	0.10	0.10	0.27	0.25	0.25	0.22
0.5	0.15	0.18	0.40	0.37	0.38	0.34
0.6	0.21	0.25	0.53	0.49	0.50	0.45
0.7	0.28	0.34	0.69	0.64	0.66	0.60
0.8	0.31	0.46	0.78	0.73	0.76	0.71
0.9	0.41	0.55	0.85	0.81	0.83	0.80
1.0	0.49	0.64	0.90	0.87	0.89	0.86
1.1	0.57	0.74	0.95	0.93	0.94	0.92
1.2	0.65	0.81	0.97	0.96	0.97	0.95
1.3	0.73	0.88	0.98	0.98	0.98	0.97
1.4	0.80	0.92	0.99	0.99	0.99	0.98
1.5	0.86	0.95	1	1	1	0.99
1.6	0.90	0.97	1	1	1	1
1.7	0.93	0.98	1	1	1	1
1.8	0.96	0.99	1	1	1	1
1.9	0.97	0.99	1	1	1	1
2.0	0.99	1	1	1	1	1

Note to Tables 2-4. The tables report empirical rejection probabilities for the nonparametric ("Nonparametric"), fluctuation ("fluctuation"), one-time reversal Sup-type ("QLR_T^*"), the $ExpW_{\infty,T}^*$ and $MeanW_T^*$ tests. The table also reports empirical rejection probabilities for a standard QLR test for breaks ("Break"). Table 2 reports results for design 1 and Table 3 for design 2 – see Section 4 for details.

Table 4. Bandwidth Selection Comparisons

	h					
β_A	0.0005	0.005	0.05	0.1	0.5	0.7
0	0.04	0.04	0.04	0.05	0.05	0.06
0.2	0.06	0.06	0.06	0.06	0.06	0.06
0.4	0.11	0.12	0.08	0.08	0.08	0.06
0.6	0.17	0.19	0.13	0.12	0.07	0.07
0.8	0.31	0.31	0.23	0.19	0.10	0.08
1.0	0.46	0.51	0.35	0.28	0.13	0.10
1.2	0.64	0.66	0.42	0.37	0.15	0.10
1.4	0.75	0.79	0.54	0.43	0.19	0.13
1.6	0.87	0.91	0.64	0.53	0.21	0.14
1.8	0.94	0.96	0.72	0.61	0.22	0.17
2.0	0.97	0.98	0.80	0.69	0.25	0.19

Note to Table 4. The table shows empirical rejection probabilities of the nonparametric test for the Monte Carlo design 1 discussed in Section 4, using different bandwidth sizes (h).

Figure 1(a)

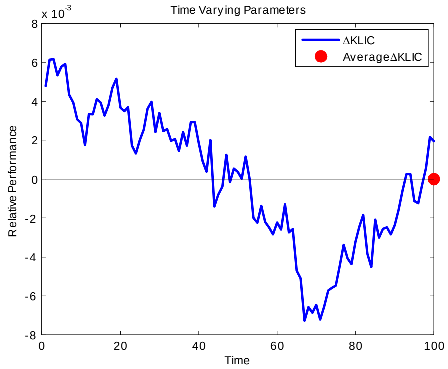


Figure 1(b)

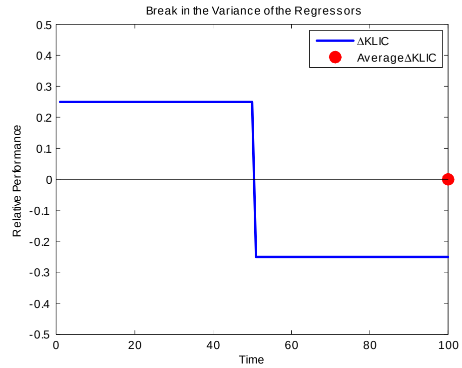


Figure 1(c)

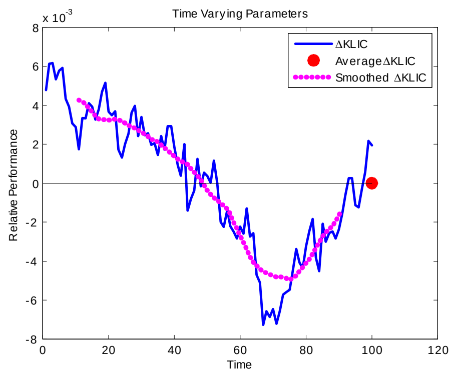


Figure 1(d)

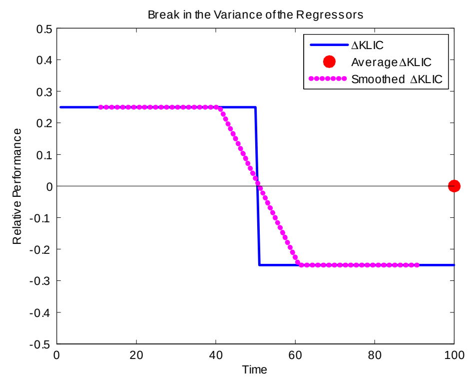


Figure 1(e)

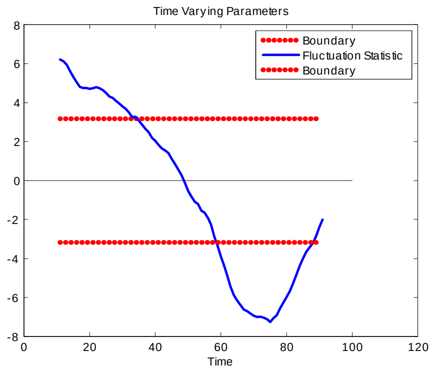


Figure 1(f)

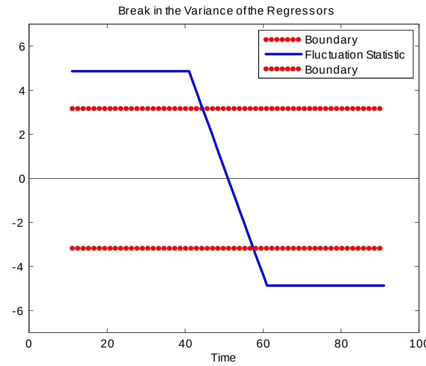


Figure 1(g)

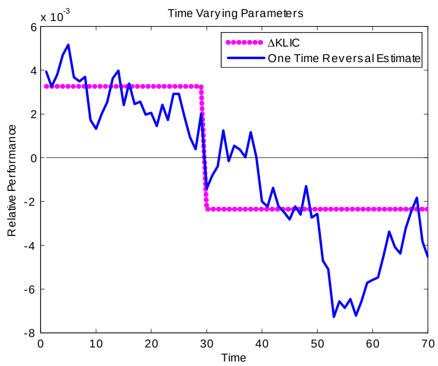


Figure 1(h)

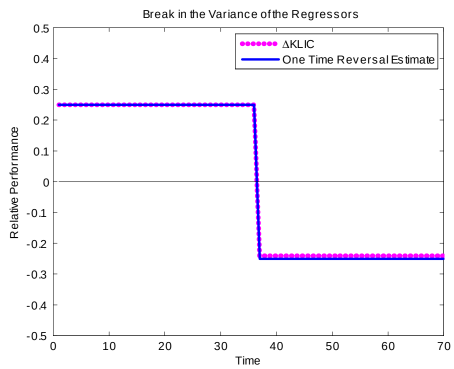
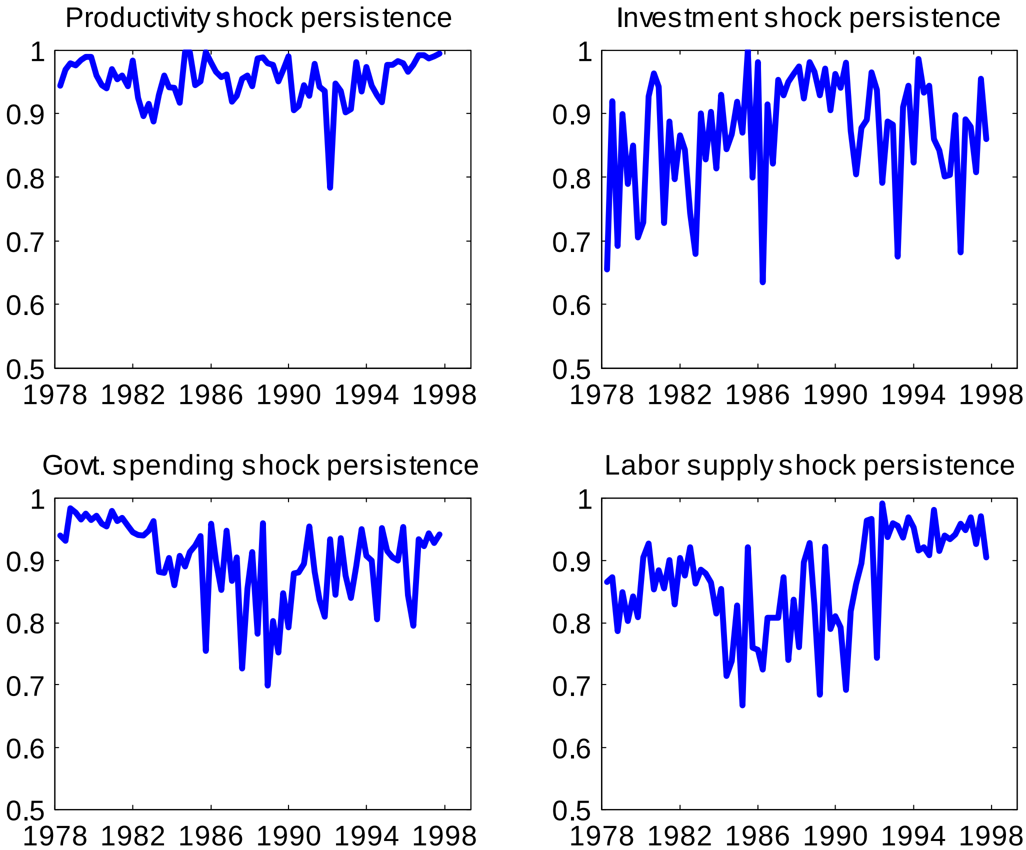
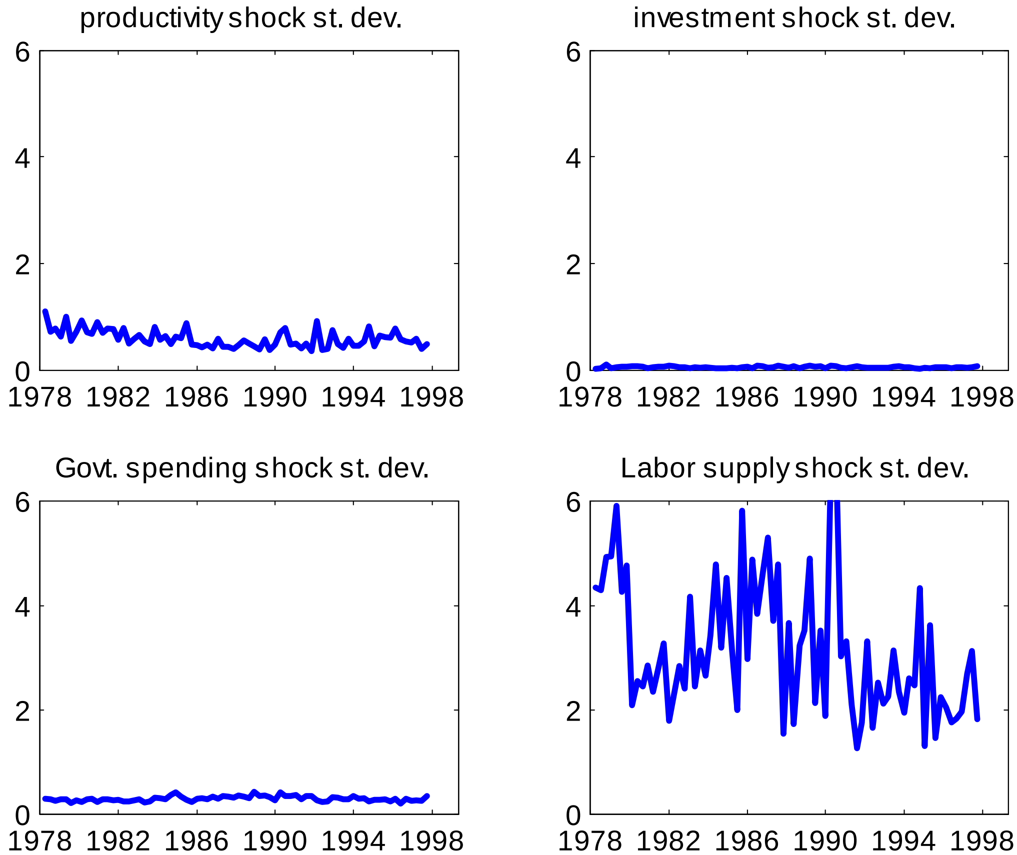


Figure 2(a). Rolling estimates of DSGE parameters (persistence of the shocks).



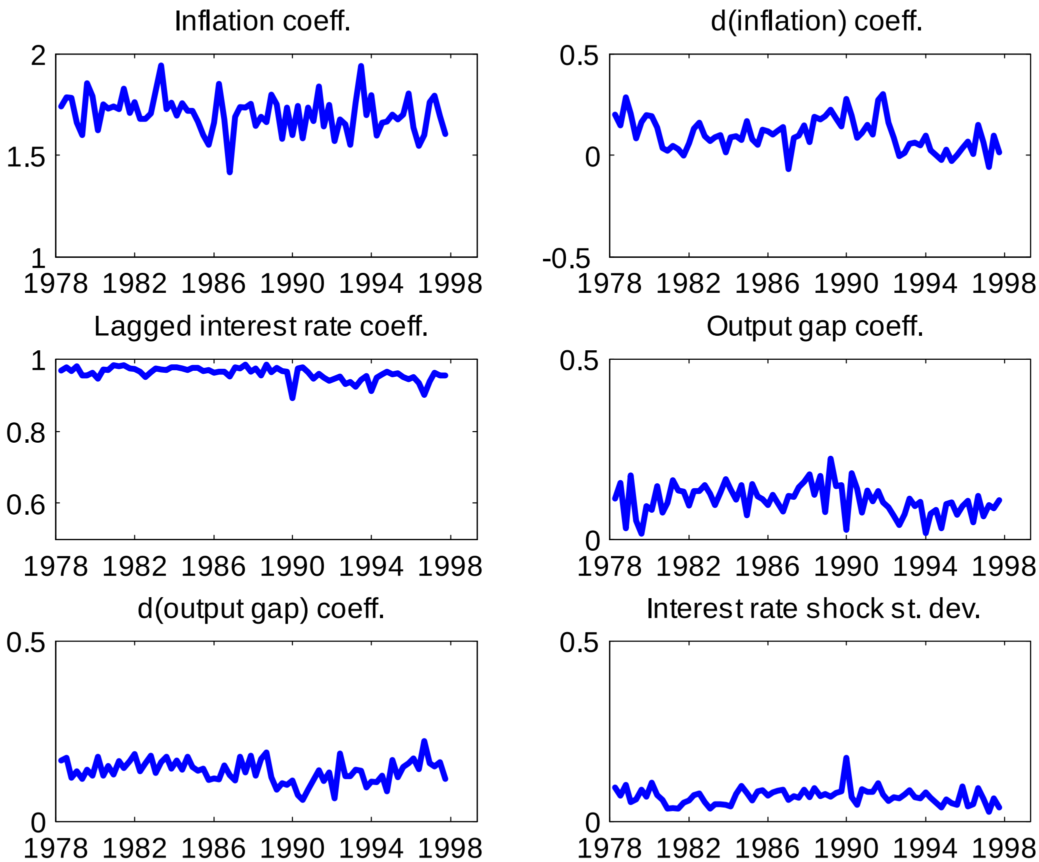
Notes to Figure 2(a). The figure plots rolling estimates of some parameters in Smets and Wouter's (2002) model. See Smets and Wouter's Table 1, p. 1142 for a description.

Figure 2(b). Rolling estimates of DSGE parameters (standard deviation of the shocks).



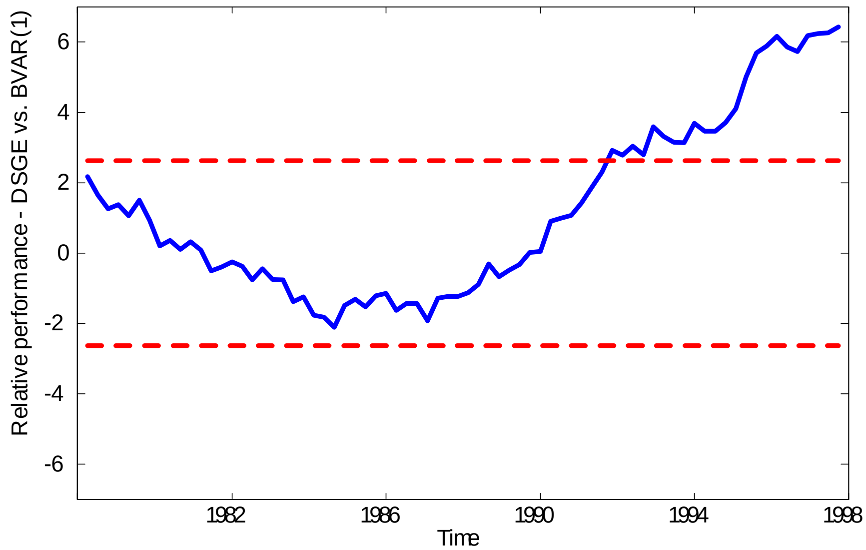
Notes to Figure 2(b). The figure plots rolling estimates of some parameters in Smets and Wouter's (2002) model using full-sample detrended data. See Smets and Wouter's Table 1, p. 1142 for a description.

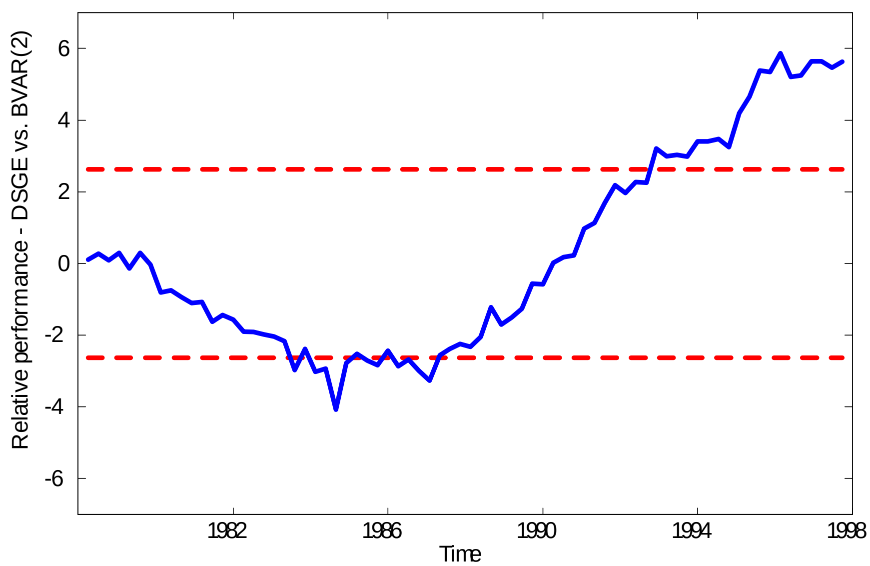
Figure 2(c). Rolling estimates of DSGE parameters (monetary policy parameters).



Notes to Figure 2(c). The figure plots rolling estimates of the parameters in the monetary policy reaction function described in Smets and Wouters' (2002) eq. (36), given by: $\widehat{R}_t = \rho \widehat{R}_{t-1} + (1-\rho)\left\{\overline{\pi}_t + r_\pi\left(\widehat{\pi}_{t-1}-\overline{\pi}_t\right) + r_Y\left(\widehat{Y}_{t-1}-\widehat{Y}_t^p\right)\right\} + r_{\Delta\pi}\left(\widehat{\pi}_t-\widehat{\pi}_{t-1}\right) + r_{\Delta Y}\left(\left(\widehat{Y}_t-\widehat{Y}_t^p\right)-\left(\widehat{Y}_{t-1}-\widehat{Y}_{t-1}^p\right)\right) + \eta_t^R$, $\overline{\pi}_t = \rho_\pi\overline{\pi}_{t-1}+\eta_t^\pi$. The figure plots: inflation coefficient (r_π), d(inflation) coefficient ($r_{\Delta\pi}$), lagged interest rate coefficient (ρ), output gap coefficient (r_Y), d(output gap) coefficient ($r_{\Delta Y}$), and standard deviation of the interest rate shock ($\sqrt{var\left(\eta_t^\pi\right)}$).

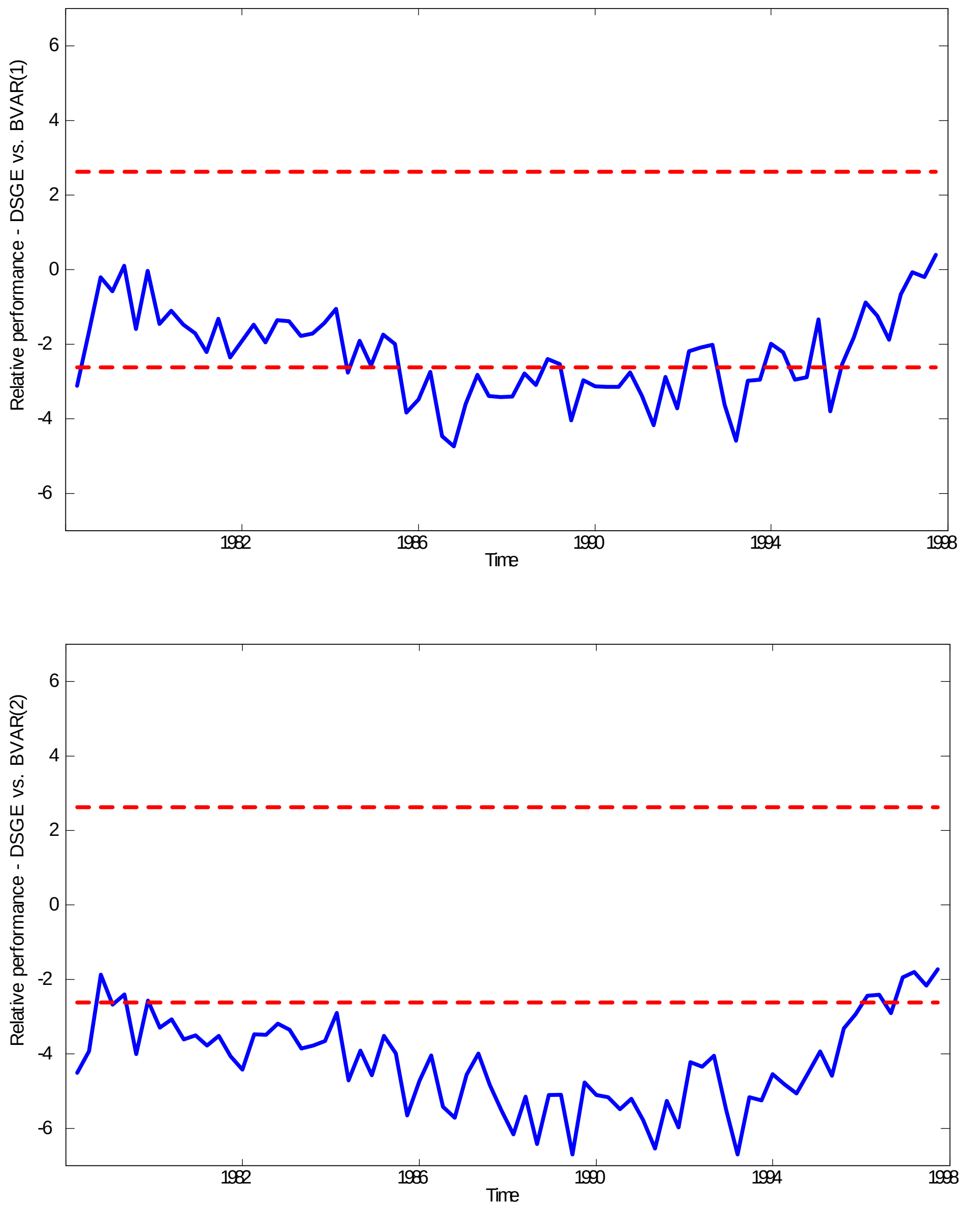
Figure 3. Fluctuation test DSGE vs. BVARs. Full-sample detrending





Notes to Figure 3. The figure plots the Fluctuation test statistic for testing equal performance of the DSGE and BVARs, using a rolling window of size $m = 70$ (the horizontal axis reports the central point of each rolling window). The 10% boundary lines are derived under the hypothesis that the local $\Delta KLIC$ equals zero at each point in time.The data is detrended by a linear trend computed over the full sample.The top panel compares the DSGE to a BVAR(1) and the lower panel compares the DSGE to a BVAR(2).

Figure 4. Fluctuation test DSGE vs. BVARs. Rolling sample detrending



Notes to Figure 4. The figure plots the Fluctuation test statistic for testing equal performance of the DSGE and BVARs, using a rolling window of size $m = 70$ (the horizontal axis reports the central point of each rolling window). The 10% boundary lines are derived under the hypothesis that the local $\Delta KLIC$ equals zero at each point in time.The data is detrended by a linear trend computed over each rolling window.The top panel compares the DSGE to a BVAR(1) and the lower panel compares the DSGE to a BVAR(2).