Model Comparisons in Unstable Environments

Raffaella Giacomini and Barbara Rossi

(*UCL and CeMMAP*) (*ICREA – Univ. Pompeu Fabra, Barcelona GSE and CREI*)

January 8, 2015

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 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 D. Kristensen for helpful comments and 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 two 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; and 2) 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 second is based on nonparametric estimates of local performance and local parameters. The fluctuation test is based on a fixed-bandwidth approximation; unreported results show that it delivers a better finite-sample performance than a test based on a standard shrinking bandwidth approximation (e.g. Wu and Zhao, 2007).

For both 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.

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 (2006), 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 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 needs 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, 2006; Muller and Petalas, 2009) in that we adapt the tools developed in that literature to our different context where the null hypothesis of interest is a joint hypothesis that the relative performance of the models is equal at each point in time and that the parameters are constant.

The fact that parameters are constant under our null hypothesis means that we are not considering the potentially relevant case in which the performance of two models is equal in spite of their parameters changing over time. The reason for excluding this case is a pragmatic one. In principle, one could have developed versions of our tests that allow for some time variation in parameters under the null hypothesis. Doing so would however be costly in terms of the general applicability of our techniques, as it would require us to impose additional restrictions on the type of time variation under the null hypothesis, the properties of the data and the models that are compatible with the assumptions on which the tests are based. We illustrate this point more concretely when discussing the assumptions of each test in the body of the paper.

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 null hypotheses and Section 4 describes the tests. Section 5 evaluates the small sample properties of our proposed procedures in a Monte Carlo experiment, and Section 6 presents the empirical results. Section 7 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 $f_t : N(\beta_t x_t, 1)$ and model 2, with density $g_t : N(\gamma_t z_t, 1)$. Here β_t and γ_t denote the pseudo-true parameters, defined as the parameters that maximize the expected log-density at time t, $\beta_t = \arg\max_\beta E[\ln f_t(\beta)]$ and $\gamma_t = \arg\max_\gamma E[\ln g_t(\gamma)]$. Even though under the assumptions considered in this example the pseudo-true parameters coincide with the true parameters, in general cases β_t and β_t^0 will be different. For instance, introducing correlation between x_t and z_t will yield $\beta_t = \beta_t^0 + \gamma_t^0 E[x_t z_t] / E[x_t^2]$.

To measure the relative distance of f_t and g_t from h_t at time t we propose using the Kullback-Leibler Information Criterion at time t, $\Delta KLIC_t$, (henceforth the "local relative KLIC"), defined as:

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

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

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

Intuitively, $\Delta KLIC_t(\theta_t)$ 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 $\beta_t^2 \sigma_{x,t}^2$. Similarly, the mis-specification of model 1 is measured by $\gamma_t^2 \sigma_{z,t}^2$. 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 the parameters change in ways that affect $\Delta KLIC_t$ differently over time, but it might also be caused by the variances of the regressors $\sigma_{x,t}^2$ and $\sigma_{z,t}^2$ changing in different ways over time while the parameters remain constant. Moreover, time-variation in pseudo-true parameters does not correspond exactly to time-variation in true parameters, as can be seen from the expression $\beta_t = \beta_t^0 + \gamma_t^0 E[x_t z_t] / E[x_t^2]$ obtained in the case of correlated regressors, as β_t could display different patterns of time variation depending on whether and how the different components change over time.

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

3 Null and Alternative Hypotheses

In this paper we construct tests of equal performance of two models over time that take into account the dependence of the relative performance on estimated parameters, and that allow for different types of time variation in both relative performance and in parameters over time. In the rest of the paper, we no longer make the distinction between true parameters and pseudo-true parameters, as all of our tests will be expressed in terms of pseudo-true parameters. For this reason, we adopt the convention of referring to pseudo-true parameters simply as parameters.

Recall that the measure of local performance is the local relative KLIC, defined as

$$\Delta KLIC_t(\theta_t) = E\left[\ln f_t\left(\beta_t\right) - \ln g_t\left(\gamma_t\right)\right], \tag{3}$$

where

$$\begin{aligned} \theta_t &= (\beta_t', \gamma_t')', \\ \beta_t &= \arg\max_b E\left[\ln f_t(b)\right], \gamma_t = \arg\max_c E\left[\ln g_t(c)\right]. \end{aligned} \tag{4}$$

We assume throughout that $\theta_t \in \Theta$, with Θ compact.

We propose two different tests, which correspond to different null and alternative hypotheses.

The first test ("one-time reversal test") assumes that under the null hypothesis the models perform equally well and the parameters are constant, whereas 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:

$$H_0^{OT} : \left\{\Delta KLIC_t\left(\theta_t\right) = 0\right\} \cap \left\{\theta_t = \theta\right\} \text{ for } t = 1, ..., T \text{ and some } \theta \in \Theta, \tag{5}$$

and

$$\begin{aligned} H_1^{OT} &: \cup_{\delta \in \Pi} \left\{\Delta KLIC_t\left(\theta_t\right) = \mu_1\left(\delta\right) 1\left(t \leq [T\delta]\right) + \mu_2\left(\delta\right) 1\left(t > [T\delta]\right)\right\} \\ \cap \left\{\theta_t &= \theta_1\left(\delta\right) 1\left(t \leq [T\delta]\right) + \theta_2\left(\delta\right) 1\left(t > [T\delta]\right)\right\}, \; t = 1, ..., T, \end{aligned} \tag{6}$$

for some $\left(\mu_1\left(\delta\right), \mu_2\left(\delta\right)\right) \neq (0, 0)$, some $\delta \in \Pi \subset (0, 1)$, and/or some $\theta_1\left(\delta\right) \neq \theta_2\left(\delta\right)$, $\theta_i\left(\delta\right) = \left(\beta_i\left(\delta\right)', \gamma_i\left(\delta\right)'\right)'$, $i = 1, 2$. Thus, $\mu_i\left(\delta\right), i = 1, 2$ are the measures of local performance and $\beta_i\left(\delta\right)$ and $\gamma_i\left(\delta\right), \; i = 1, 2$ are the local parameters for the sub-samples before and after the reversal, which occurs at the unknown fraction of the total sample δ. The one-time reversal test thus 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, $[T\delta]$, can be estimated and the path of relative performance equals $\mu_1\left(\delta\right)$ before the change and $\mu_2\left(\delta\right)$ after the change.

The second test ("fluctuation test") is based on a nonparametric estimator of relative performance. The fluctuation test uses a novel asymptotic approximation which assumes that the bandwidth is fixed. In this approximation, consistent estimation of the local relative performance $\Delta KLIC_t(\theta_t)$ 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, computed at the smoothed local parameter:

$$\text{Smoothed local relative KLIC} : \Delta KLIC_t^*(\theta_t^*) = E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\left(\ln f_j(\beta_t^*) - \ln g_j\left(\gamma_t^*\right)\right)\right], \tag{7}$$

where $\theta_t^* = (\beta_t^*, \gamma_t^*)'$ is defined as

$$\beta_t^* = \arg\max_b E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\ln f_j(b)\right], \tag{8}$$

(and similarly for γ_t^*), with $K(\cdot)$ a kernel function and h the bandwidth.

The fluctuation test corresponds to different null and alternative hypotheses:

$$H_0^{FB} : \{\Delta KLIC_t^*(\theta_t^*) = 0\} \cap \{\theta_t^* = \theta^*\} \text{ for } t = 1, ..., T \text{ and some } \theta^* \in \Theta, \tag{9}$$

and

$$H_1^{FB} : \Delta KLIC_t^*(\theta_t^*) \neq 0 \text{ at some } 1 \leq t \leq T. \tag{10}$$

In the example in Section 2, using a rectangular kernel with bandwidth $h = m/T$ (and assuming for simplicity that m is an even number) we can see that $\Delta KLIC_t^*(\theta_t^*)$ differs from $\Delta KLIC_t(\theta_t)$ in (2), since, first of all, $\beta_t^* = \left(\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\sigma_{x,j}^2\beta_j^0\right) / \left(\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\sigma_{x,j}^2\right) \neq \beta_t^0$, whereas $\beta_t = \beta_t^0$, and

$$\Delta KLIC_t^*(\theta_t^*) = \frac{1}{2}\left[\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left[2\beta_t^0\beta_t^* - (\beta_t^*)^2\right]\sigma_{x,j}^2 - \frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\left[2\gamma_t^0\gamma_t^* - (\gamma_t^*)^2\right]\sigma_{z,j}^2\right]. \tag{11}$$

An important point to emphasize is that for both tests the null hypothesis is a joint hypothesis of equal performance and constant parameters. This in practice rules out situations in which two models have equal performance but their parameters are changing over time. The assumption of constant parameters under the null hypothesis is in principle stronger than necessary, but it facilitates the statement and the verification of the assumptions on which the tests rely. For example, allowing for time-varying parameters under the null hypothesis would be difficult to reconcile with the assumption of constant variance of the loss differences that we make for all tests, since in general the variance of the loss differences will depend on the models' parameters. Both assumptions

of constant parameters and constant variance could be relaxed in the context of specific models and/or for a specific test. For example, one could allow for time variation in parameters that disappears asymptotically, or make local stationarity assumptions such as those considered in the nonparametric estimation literature (e.g., Kristensen, 2013). One could also relax the constant variance assumption and rely on bootstrap methods to derive the tests, along the lines of Cavaliere and Taylor (2005).

The difference between the various alternative hypotheses as well as the difference between $\Delta KLIC_t(\theta_t)$ and $\Delta KLIC_t^*(\theta_t^*)$ 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 varying smoothly as a random walk whereas γ_t, $\sigma^2_{x,t}$, $\sigma^2_{z,t}$ are constant, $t = 1, ..., 100$. In the second scenario (right panels of Figure 1), $\beta_t, \gamma_t, \sigma^2_{z,t}$ are constant but the relative performance is time-varying because $\sigma^2_{x,t}$ has a break at $T/2$.

INSERT FIGURE 1 HERE

Figures 1(a) and 1(b) report the local relative $KLIC_t$ in equation (1) in the two scenarios. Figures 1(c) and 1(d) show $\Delta KLIC_t$ as well as $\Delta KLIC_t^*$ in equation (7) computed using a bandwidth $m/T = 1/5$. 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 (9) and (5). One can see that the 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 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 two statistical tests. The one-time reversal test of hypothesis (5) 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 fluctuation test of hypothesis (9) relies on constructing confidence bands for the different object $\Delta KLIC_t^*$ in (7) under the null hypothesis by using a 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 parameters.

The two 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 one-time reversal test (Figure 1(f)) will in principle capture it; conversely, the fluctuation test (Figure 1(d)), which relies on the smoothed relative KLIC, will smooth out the time variation: the time variation will thus be more difficult to detect, lowering the power of the test (again, depending on the choice of bandwidth). 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 test 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(c).

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 models' smoothed local relative $KLIC$ with a pre-specified probability level under the null hypothesis that the relative performance 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 local average $\Delta KLIC$ to ascertain which model performed best at any point in time.

Figures 1(g,h) illustrate the one-time reversal test[2] 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 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.

[2]The One-time Reversal test is implemented as a Sup-type test. See Section 4.1 for more details.

4 Tests of Stability in the Relative Performance of Models

In this section, we derive the two classes of tests assessing the stability in the relative performance of two models over time.

Each test assumes 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.

4.1 The One-time Reversal Test

The first test is a test of the null hypothesis (5) against (6), which draws from the literature on testing the stability of the mean of a time series (e.g. Andrews, 1993). These tests 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.

The test is implemented as follows. For a given $\delta \in \Pi \subset (0,1)$, let $\widehat{\mu}(\delta) \equiv [\widehat{\mu}_1(\delta),\ \widehat{\mu}_2(\delta)]$, where:

$$\widehat{\mu}_1(\delta) = \frac{1}{[T\delta]} \sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\theta}_1(\delta)\right),\ \widehat{\mu}_2(\delta) = \frac{1}{[T(1-\delta)]} \sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\theta}_2(\delta)\right) \tag{12}$$

and $\widehat{\theta}_1(\delta) = (\widehat{\beta}_1(\delta)', \widehat{\gamma}_1(\delta)')'$, $\widehat{\theta}_2(\delta) = (\widehat{\beta}_2(\delta)', \widehat{\gamma}_2(\delta)')'$, $\Delta L_t(.) = \ln f_t(.) - \ln g_t(.)$, with

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

(and similarly for $\widehat{\gamma}_1(\delta), \widehat{\gamma}_2(\delta)$). Also, let $\widehat{\beta}_T = \arg\max_b \left(\frac{1}{T} \sum_{t=1}^{T} \ln f_t(b) \right)$ (and similarly for $\widehat{\gamma}_T$), and $\widehat{\theta}_T \equiv \left[\widehat{\beta}_T', \widehat{\gamma}_T'\right]'$.

The test relies on the following assumptions:

Assumptions OT: Let θ be the constant value of θ_t under the null hypothesis. The following holds: (1) $\left\{T^{-1/2} \sum_{j=1}^{[\lambda T]} \Delta L_j(\theta)\right\}$ obeys a Functional Central Limit Theorem (FCLT) under $H_{0,T}^{OT}$ for $0 \leq \lambda \leq 1$, such that: $\sigma^{-1} T^{-1/2} \sum_{j=1}^{[\lambda T]} \Delta L_j(\theta) \Longrightarrow \mathcal{B}(\lambda)$ (a standard Brownian Motion process); (2) $sup_{\delta \in \Pi} ||\widehat{\beta}_1(\delta) - \beta|| = o_p(1)$ and $sup_{\delta \in \Pi} ||T^{1/2}\left(\widehat{\beta}_1(\delta) - \beta\right)|| = O_p(1)$ under $H_{0,T}^{OT}$ as $T \to \infty$ (and similarly for $\widehat{\gamma}_1(\delta)$, $\widehat{\beta}_2(\delta)$, $\widehat{\gamma}_2(\delta)$, $\widehat{\beta}_T$, $\widehat{\gamma}_T$); (3) (a) the log-likelihoods of both models, $\ln f_T(\beta, \delta) = \sum_{t=1}^{T} \ln f_t(\beta, \delta)$ and $\ln g_T(\gamma, \delta) = \sum_{t=1}^{T} \ln g_t(\gamma, \delta)$, do not depend on δ for all β, γ in the

null hypothesis; (b) θ *is an interior point of the parameter space* Θ*; (c)* $f_T\left(\widetilde{\beta},\delta\right), g_T\left(\widetilde{\gamma},\delta\right)$ *are twice continuously partially differentiable in* $\widetilde{\beta},\widetilde{\gamma}$ *for all* $\delta\in\Pi$ *and* $\widetilde{\beta},\widetilde{\gamma}$ *in some neighborhood of the null,* Θ_0*; (d)* $-B_T^{-1}\nabla^2\ln f_T\left(\widetilde{\beta},\delta\right)B_T^{-1}\underset{p}{\to}\Phi_\beta\left(\widetilde{\beta},\delta\right), -B_T^{-1}\nabla^2\ln g_T\left(\widetilde{\gamma},\delta\right)B_T^{-1}\underset{p}{\to}\Phi_\gamma\left(\widetilde{\gamma},\delta\right)$ *uniformly over* $\delta\in\Pi$ *and* $\widetilde{\beta},\widetilde{\gamma}\in\Theta_0$ *under* β,γ *for some nonrandom matrix functions* $\Phi_\beta\left(\beta,\delta\right),\Phi_\gamma\left(\gamma,\delta\right)$ *and some sequence of nonrandom diagonal matrices* $\{B_T : T\geq 1\}$ *whose elements diverge to infinity as* $T\to\infty$*;*[3] *(e)* $\Phi_\beta\left(\widetilde{\beta},\delta\right),\Phi_\gamma\left(\widetilde{\gamma},\delta\right)$ *are uniformly continuous in* $\left(\widetilde{\theta},\delta\right)$ *over* $\Theta_0\times\Pi$*; (f)* $\Phi_\beta\left(\beta,\delta\right),\Phi_\gamma\left(\gamma,\delta\right)$ *are uniformly positive definite over* $\delta\in\Pi$*; (4)* $\lim_{T\to\infty} var\left(T^{-1/2}\sum_{t=1}^{T}\Delta L_t\left(\theta\right)\right) = \sigma^2>0$ *is constant and finite, and* $\widehat{\sigma}^2$ *is a consistent estimator of* σ^2*.*

Assumption OT(1) assumes a FCLT for partial sum processes. Assumptions OT(2,3) are standard assumptions that guarantee that the estimated parameters as well as the score functions obey regularity conditions ensuring their convergence. In particular, Assumption OT(3) follow from assumptions similar to Andrews and Ploberger (1994). Assumption OT(4) imposes that the variance is constant under the null hypothesis, and a consistent estimator under the null hypothesis is for example a standard HAC estimator

$$\widehat{\sigma}^2 = \sum_{i=-q(T)+1}^{q(T)-1}\left(1-|i/q(T)|\right)T^{-1}\sum_{j=q(T)}^{T+1-q(T)}\Delta L_j^d\left(\widehat{\theta}_T\right)\Delta L_{j-i}^d\left(\widehat{\theta}_T\right), \tag{13}$$

where $q(T)$ is a bandwidth that grows with T (e.g., Newey and West, 1987, Andrews, 1991), and $\Delta L_j^d\left(.\right)$ indicates demeaned $\Delta L_j^d\left(.\right)$.

Under Assumption OT, we provide Sup-type tests for the one-time reversal in the following proposition:[4]

Proposition 1 (Sup-type Test) *Suppose Assumption OT holds. Let* $QLR_T^* = \sup_{\delta\in\Pi}\Phi_T\left(\delta\right)$, $\Phi_T\left(\delta\right) = LM_1 + LM_2\left(\delta\right)$*, where* $\Pi\subset(0,1)$ *and*

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

Under the null hypothesis H_0^{OT}*, we have:* $QLR_T^* \Longrightarrow \sup_{\delta\in\Pi}\left[\frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)} + \mathcal{B}\left(1\right)^2\right]$*, and* $\mathcal{B}\left(\cdot\right)$ *and* $\mathcal{BB}\left(\delta\right)\equiv\mathcal{B}\left(\delta\right)-\delta\mathcal{B}\left(1\right)$ *are, respectively, a standard univariate Brownian motion and a Brownian bridge. The null hypothesis is thus rejected when* $QLR_T^* > k_\alpha$*. The critical values* (α, k_α) *are:* $(0.05, 9.8257)$, $(0.10, 8.1379)$.

[3] ∇^2 denotes the second derivative with respect to the parameter.

[4] Sup-type tests have been used in the parameter instability literature since Andrews (1993). Note that the sup-type test could alternatively be implemented as $\sup_{\delta\in\Pi}W_T\left(\delta\right)$, where $W_T\left(\delta\right)$ is defined in eq. (14).

Among the advantages of the Sup-type 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 to estimate the time of the switch in the relative performance; (iii) the test is designed against one time breaks in the relative performance. Here below is a step by step procedure to implement the approach suggested in Proposition 1 with an overall significance level α:

(i) test the hypothesis of equal performance at each time by using the statistic QLR_T^* from Proposition 1 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.[5]

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

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

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

We also provide tests similar in spirit to those proposed by Andrews and Ploberger (1994). The tests rely on the Wald-type test statistic, rather than LM-type.[6]

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

$$W_T(\delta) = T\widehat{\mu}(\delta)' H' \left(H\mathcal{I}_{T,\delta}^{-1}H'\right)^{-1} H\widehat{\mu}(\delta) \tag{14}$$

$$ExpW_{\infty,T}^* = \ln\frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0}\exp\left(\frac{1}{2}W_T(\delta)\right)d\delta; \tag{15}$$

$$MeanW_T^* = \frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0} W_T(\delta)\,d\delta, \tag{16}$$

[5] 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 distribution whereas those for LM_2 are from Andrews (1993).

[6] The Wald-type test allows for a more general variance estimator. One could also implement the Sup-type test in Proposition 1 as $QLR_T^* = \sup_{\delta\in\Pi} W_T(\delta)$.

where $\delta_0 = 0.15$, $H \equiv \begin{pmatrix} 1 & -1 \\ \delta & 1-\delta \end{pmatrix}$, $\mathcal{I}_{T,\delta}^{-1} = \begin{pmatrix} \delta^{-1}\widehat{\sigma}_1^2 & 0 \\ 0 & (1-\delta)^{-1}\widehat{\sigma}_2^2 \end{pmatrix}$, $\widehat{\sigma}_1^2$ *is a HAC estimator of the asymptotic variance of* $\Delta L_t\left(\widehat{\theta}_1(\delta)\right)$, $t = 1, ..., [T\delta]$ *and* $\widehat{\sigma}_2^2$ *is a HAC estimator of the variance of* $\Delta L_t\left(\widehat{\theta}_2(\delta)\right)$, $t = [T\delta]+1, ..., T$.[7] *Under the null hypothesis* H_0^{OT},

$$ExpW_{\infty,T}^* \implies \ln\left[\frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0}\exp\left(\frac{1}{2}\frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)}+\frac{1}{2}\mathcal{B}(1)^2\right)d\delta\right], \tag{17}$$

$$MeanW_T^* \implies \frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0}\left(\frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)}+\mathcal{B}(1)^2\right)d\delta, \tag{18}$$

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* $(\alpha; \kappa_\alpha, v_\alpha)$ *are:* $(0.05; 3.13, 5.36)$ *and* $(0.10; 2.44, 4.26)$.

The power properties of these tests will be evaluated in Section 5.

4.2 The Fluctuation Test

In this section, we derive a test of the hypothesis (9) against the alternative (10). The test relies on constructing a nonparametric estimate of the local relative performance; however we consider an asymptotic approximation in which the bandwidth is fixed instead of the common approximation used in the nonparametric literature which requires that the bandwidth shrinks as the sample grows. For simplicity, we restrict attention to a rectangular kernel, but the analysis could be easily extended to the case of a general kernel. For a particular choice of fixed bandwidth $m = [hT]$, we thus define the smoothed local relative KLIC using a rectangular kernel as:

$$\Delta KLIC_t^*(\theta_t^*) = m^{-1}\sum_{j=t-m/2+1}^{t+m/2} E[\Delta L_j(\theta_t^*)], \quad t = m/2, ..., T-m/2. \tag{19}$$

Therefore, the null and alternative hypotheses become:

$$H_0^{FB} : \{\Delta KLIC_t^*(\theta_t^*) = 0\} \cap \{\theta_t^* = \theta^*\} \text{ for } t = m/2, ..., T-m/2 \text{ and some } \theta^* \in \Theta \tag{20}$$

$$\text{against} \tag{21}$$

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

We estimate the smoothed local relative KLIC in eq. (19) as:

$$m^{-1}\sum_{j=t-m/2+1}^{t+m/2}\Delta L_t\left(\widehat{\theta}_t^*\right), \tag{22}$$

[7] The formula is similar to eq. (13) except that it applies to the relevant sub-sample.

where $\widehat{\theta}_t^* = \left[\widehat{\beta}_t^{*\prime}, \widehat{\gamma}_t^{*\prime}\right]'$ are nonparametric estimates of the local parameters obtained as the solution to (e.g., for the first model)

$$m^{-1} \sum_{j=t-m/2+1}^{t+m/2} \nabla \ln f_t\left(\widehat{\beta}_t^*\right) = 0, \tag{23}$$

with $\nabla \ln f_t(\cdot)$ denoting the first derivative of the log-likelihood at time t. The fact that under the null hypothesis the parameters are constant means that one could in principle obtain a valid test by letting the measure of local performance (22) depend on any other estimator of the parameters, and not necessarily a local estimator. The reason for considering a local estimator of the parameters is to obtain a test that has power against smooth time variation in parameters under the alternative hypothesis (although we do not formally derive the properties of the test under the alternative hypothesis, but only those under the null hypothesis).

Our proposed test, which we call the fluctuation test, can be derived under the following assumptions:[8]

Assumptions FB: Let θ^* be the constant value of θ_t^* under the null hypothesis and $t = [\lambda T]$. The following holds: (1) $\left\{T^{-1/2}\sum_{j=1}^{[\lambda T]} \Delta L_j(\theta^*)\right\}$ obeys a Functional Central Limit Theorem (FCLT) under $H_{0,T}^{FB}$ for $\lambda \in \Pi \subset (0,1)$, such that: $\sigma^{-1}T^{-1/2}\sum_{j=1}^{[\lambda T]} \Delta L_j(\theta^*) \Longrightarrow \mathcal{B}(\lambda)$ (a standard Brownian Motion process); (2) $sup_\lambda||\widehat{\beta}_{[T\lambda]}^* - \beta^*|| = o_p(1)$ and $sup_\lambda||T^{1/2}\left(\widehat{\beta}_{[T\lambda]}^* - \beta^*\right)|| = O_p(1)$ under $H_{0,T}^{FB}$ as $T \to \infty$ (and similarly $\widehat{\gamma}_{[T\lambda]}^*(\delta)$); (3) $sup_\lambda||m^{-1}\sum_{j=[T\lambda]-[T\mu]/2}^{[T\lambda]+[T\mu]/2} \nabla^2 \ln f_j\left(\ddot{\beta}_{[T\lambda]}\right) - E\left(\nabla^2 \ln f_j(\beta^*)\right)|| \underset{p}{\to} 0$ whenever $\ddot{\beta}_{[T\lambda]}$ satisfies $sup_\lambda||\ddot{\beta}_{[T\lambda]} - \beta^*|| \underset{p}{\to} 0$ as $m, T \to \infty$, and $E\left(\nabla^2 \ln f_j(\beta^*)\right)$ is positive and finite; (4) $\lim_{T\to\infty} var\left(T^{-1/2}\sum_{t=1}^{T} \Delta L_t(\theta^*)\right) = \sigma^2 > 0$ is constant and finite, and $\widehat{\sigma}^2$ is a consistent estimator of σ^2; (5) $m/T = h$, with $h \in (0,\infty)$ and $m, T \to \infty$.

The assumptions underlying the fluctuation test are similar to those considered for the one-time reversal test. In particular we require the loss differences to satisfy a FCLT when evaluated at the pseudo-true parameters, which are assumed to be constant under the null hypothesis. We also assume the asymptotic variance to be constant under the null hypothesis, which again is a stronger requirement than necessary, but it facilitates the statement of the FCLT. Assumptions FB(2,3) are high-level but standard; more primitive conditions can be specified in the context of specific models. The main difference between assumption FB and assumptions used in the non-parametric literature is that the fluctuation test considers a bandwidth that is a fixed proportion of the total sample size (assumption (5)).

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

One can verify that Assumption FB(1) is satisfied in the example of Section 2, where $\Delta L_t(\theta^*) = \frac{1}{2}\left\{\left[2\beta_t^0\beta^* - (\beta^*)^2\right]x_t^2 - \left[2\gamma_t^0\gamma^* - (\gamma^*)^2\right]z_t^2\right\}$. Since x_t and z_t are i.i.d., under the further assumption that the true parameters β_t^0 and γ_t^0 are also constant under the null hypothesis, $\Delta L_t(\theta^*)$ is i.i.d. and σ_t^2 is constant, which satisfies the assumptions of Donsker's FCLT theorem. It is easy to verify that Assumptions FB (2,3) hold in the linear model case that we consider.

The following proposition provides a justification for the fluctuation test.

Theorem 3 (Fluctuation Test) *Suppose Assumption FB holds. Consider the test statistic*

$$\max_{t=m/2,\ldots,T-m/2}|F_t| = \max_{t=m/2,\ldots,T-m/2}\left|\widehat{\sigma}^{-1}m^{-1/2}\sum_{j=t-m/2+1}^{t+m/2}\Delta L_j(\widehat{\theta}_t^*)\right|, \tag{24}$$

where $\widehat{\theta}_t^*$ is as in discussed in equation (23) and $\widehat{\sigma}^2$ is a consistent estimator of σ^2, e.g. as in equation (13). Under the null hypothesis (20)

$$F_t \Longrightarrow \left[\mathcal{B}(\lambda + h/2) - \mathcal{B}(\lambda - h/2)\right]/\sqrt{h}, \tag{25}$$

where $t = [\lambda T]$, $h = m/T$ 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

$$\Pr\left\{\max_{\lambda}\left|\left[\mathcal{B}(\lambda + h/2) - \mathcal{B}(\lambda - h/2)\right]/\sqrt{h}\right| > k_\alpha\right\} = \alpha. \tag{26}$$

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 $h = m/T$.

INSERT TABLE 1 HERE

The fluctuation test relies on a specific choice of bandwidth h, and the results of the test will be different for different choices of h. As a practical recommendation, we suggest assessing the sensitivity of the test to a few different choices of h, which is easy to do as we only provide critical values for several possible choices of h.

5 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 $\Delta KLIC_t$ 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:

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

where $x_t \sim N\left(0,\sigma_{x,t}^2\right)$, $z_t \sim N\left(0,\sigma_{z,t}^2\right)$, $t = 1,2,...,T$, $T = 200$. 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. $\sigma_{x,t}^2 = \sigma_{z,t}^2 = 1$, $\gamma_t = 1$, $\beta_t = 1 + \beta_A \cdot 1\left(t \leq 0.5T\right) - \beta_A \cdot 1\left(t > 0.5T\right)$. In this design, we let the parameter β change over time, and this affects the relative performance of the models over time.

Design 2. $\sigma_{x,t}^2 = 1 + \sigma_A^2 \cdot 1\left(t > 0.75T\right)$, $\sigma_{z,t}^2 = 1$, $\beta_t = 1$, $\gamma_t = 1$. In this design, the parameters in the conditional mean are constant but one of the variances ($\sigma_{x,t}^2$) 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%. The Fluctuation test is implemented with $h = 0.3$. 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. 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 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

6 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 a-theoretical Bayesian VARs (BVARs). Furthermore, they find 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 a-theoretical 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, (the local $\Delta KLIC$), 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).[9] We present results for two different transformations of the data. The first applies the same de-trending 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 de-trending"). 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 de-trending is performed on each rolling estimation window ("rolling-sample de-trending").

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

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 de-trending of the data. The estimate of the local relative $KLIC$ is

[9] 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.

evaluated at the posterior modes of the models' parameters computed over the rolling windows, using the fact that they 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 de-trended 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 de-trending 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).

7 Conclusions

This paper developed statistical testing procedures for evaluating the relative performance of two competing models in unstable environments. We proposed two tests: 1) a one-time reversal test; and 2) a fluctuation test. We investigated the advantages and limitations of the two approaches and compared the quality of the approximation that they deliver in finite samples. Based on the results of the latter, 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 de-trending 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] 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.

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

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

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

[8] Giacomini, R. and B. Rossi (2010), "Forecast Comparisons in Unstable Environments", *Journal of Applied Econometrics* 25(4), 595-620.

[9] Kristensen, D. (2013), "Nonparametric Estimation of Time-Varying Parameters", mimeo

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

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

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

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

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

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

[16] 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.

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

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

8 Appendix A - Proofs

Lemma 4 *Let* $X_T(\delta) = o_{p,\delta}(1)$ *denote* $sup_{\delta\in\Pi}||X_T(\delta)|| = o_p(1)$ *and* $X_T(\delta) = O_{p,\delta}(1)$ *denote* $sup_{\delta\in\Pi}||X_T(\delta)|| = O_p(1)$. *Under Assumption OT and* H_0^{OT},

$$T^{-1/2}\sum_{t=1}^{[T\delta]}\left(\ln f_t(\widehat{\beta}_1(\delta)) - \ln g_t(\widehat{\gamma}_1(\delta))\right) = T^{-1/2}\sum_{t=1}^{[T\delta]}(\ln f_t(\beta(\delta)) - \ln g_t(\gamma(\delta))) + o_{p,\delta}(1).$$

Proof of Lemma (4). By applying a second order Taylor expansion around $\theta = \left[\beta', \gamma'\right]'$:

$$\begin{aligned}
& T^{-1/2}\sum_{t=1}^{[T\delta]}\left(\ln f_t(\widehat{\beta}_1(\delta)) - \ln g_t(\widehat{\gamma}_1(\delta))\right) \\
= \; & T^{-1/2}\sum_{t=1}^{[T\delta]}(\ln f_t(\beta) - \ln g_t(\gamma)) + \\
& + \frac{1}{2}T^{-1}\sum_{t=1}^{[T\delta]}\nabla \ln f_t(\widehat{\beta}_1(\delta))\left(\widehat{\beta}_1(\delta) - \beta\right)T^{1/2} && (27)\\
& - \frac{1}{2}T^{-1}\sum_{t=1}^{[T\delta]}\nabla \ln g_t(\widehat{\gamma}_1(\delta))(\widehat{\gamma}_1(\delta) - \gamma)T^{1/2} && (28)\\
& + \frac{1}{2}\left(\widehat{\beta}_1(\delta) - \beta\right)'\left[T^{-1}\sum_{t=1}^{[T\delta]}\nabla^2 \ln f_t(\ddot{\beta}_{1,T}(\delta))\right]\left(\widehat{\beta}_1(\delta) - \beta\right)T^{1/2} && (29)\\
& - \frac{1}{2}(\widehat{\gamma}_1(\delta) - \gamma)'\left[T^{-1}\sum_{t=1}^{[T\delta]}\nabla^2 \ln g_t\left(\ddot{\gamma}_{1,T}(\delta)\right)\right](\widehat{\gamma}_1(\delta) - \gamma)T^{1/2} && (30)\\
= \; & T^{-1/2}\sum_{t=1}^{[T\delta]}(\ln f_t(\beta) - \ln g_t(\gamma)) + o_{p,\delta}(1) + o_{p,\delta}(1), && (31)
\end{aligned}$$

where $\ddot{\beta}_{1,T}(\delta)$ is an intermediate point between $\widehat{\beta}_1(\delta)$ and β (similarly for $\ddot{\gamma}_{1,T}(\delta)$). By definition of the ML estimator, $\sum_{t=1}^{[T\delta]}\nabla \ln f_t(\widehat{\beta}_1(\delta)) = 0$ (similarly for $\nabla \ln g_t$). By Assumption OT(2), $T^{1/2}\left(\widehat{\beta}_1(\delta) - \beta\right) = O_{p,\delta}(1)$ which proves that (27) and (28) are $o_{p,\delta}(1)$. Furthermore, Assumptions OT(2,3) ensure that, under the null hypothesis, $T^{-1}\sum_{t=1}^{[T\delta]}\nabla^2 \ln f_t(\ddot{\beta}_{1,T}(\delta)) = O_{p,\delta}(1)$ (and similarly for the component in γ); by Assumption OT(2), $T^{1/2}\left(\widehat{\beta}_1(\delta) - \beta\right) = O_{p,\delta}(1)$ and $\left(\widehat{\beta}_1(\delta) - \beta\right) = o_{p,\delta}(1)$ (and similarly for the components in γ), proving that (29) and (30) are $o_{p,\delta}(1)$. ■

Proof of Proposition 1. By Lemma 4 and similar arguments, as well as the consistency of $\widehat{\sigma}$, we have:

$$(i)\; LM_1 = \sigma^{-2}\left[T^{-1/2}\sum_{t=1}^{T}(\ln f_t(\beta_t) - \ln g_t(\gamma_t))\right]^2 + o_{p,\delta}(1)$$

and

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

By Assumption OT, under the null hypothesis:

$$\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}(\ln f_t(\beta_t)-\ln g_t(\gamma_t)) \Longrightarrow \mathcal{B}(1) \tag{32}$$

$$\begin{aligned}&\sigma^{-1}\delta^{-1/2}(1-\delta)^{-1/2}\left[T^{-1/2}\sum_{t=1}^{[T\delta]}(\ln f_t(\beta_t)-\ln g_t(\gamma_t))\right. \\ &\left.-\delta T^{-1/2}\sum_{t=1}^{T}(\ln f_t(\beta_t)-\ln g_t(\gamma_t))\right] \\ \Longrightarrow\ &\delta^{-1/2}(1-\delta)^{-1/2}[\mathcal{B}(\delta)-\delta\mathcal{B}(1)] = \delta^{-1/2}(1-\delta)^{-1/2}\mathcal{BB}(\delta),\end{aligned} \tag{33}$$

where the limiting distributions in (32) and (33) are asymptotically uncorrelated.[10] Then:

$$\begin{aligned}LM_1+LM_2(\delta) &= \sigma^{-2}\left[T^{-1/2}\sum_{t=1}^{T}(\ln f_t(\beta_t)-\ln g_t(\gamma_t))\right]^2 \\ &\quad +\sigma^{-2}\delta^{-1}(1-\delta)^{-1}\left[T^{-1/2}\sum_{t=1}^{[T\delta]}(\ln f_t(\beta_t)-\ln g_t(\gamma_t))\right. \\ &\quad \left.-\delta T^{-1/2}\sum_{t=1}^{T}(\ln f_t(\beta_t)-\ln g_t(\gamma_t))\right]^2+o_{p,\delta}(1) \\ \Longrightarrow\ &\mathcal{B}(1)^2+\delta^{-1}(1-\delta)^{-1}\mathcal{BB}(\delta)^2\end{aligned} \tag{34}$$

and the result follows by the Continuous Mapping Theorem. ■

Proof of Corollary 2. The proof follows from $\sigma^{-1}T^{1/2}\widehat{\mu}(\delta) \Longrightarrow \left(\delta^{-1}\mathcal{B}(\delta),(1-\delta)^{-1}[\mathcal{B}(1)-\mathcal{B}(\delta)]\right)'$ using Lemma 4 and arguments similar to those in Proposition 1. Thus, $\sigma^{-1}HT^{1/2}\widehat{\mu}(\delta) \Longrightarrow \left[\frac{\mathcal{BB}(\delta)}{\delta(1-\delta)},\mathcal{B}(1)\right]'$, which implies $W_T(\delta) \Longrightarrow \frac{\mathcal{BB}(\delta)^2}{\delta(1-\delta)}+\mathcal{B}(1)^2$, and the result obtains by applying the Continuous Mapping Theorem. ■

[10] See Rossi (2005).

Proof of Theorem 3. Let $X_T(\lambda) = o_{p,\lambda}(1)$ denote $sup_\lambda ||X_T(\lambda)|| = o_p(1)$ and $X_T(\lambda) = O_{p,\lambda}(1)$ denote $sup_\lambda ||X_T(\lambda)|| = O_p(1)$. Let $\sum_j \equiv \sum_{j=t-m/2+1}^{t+m/2}$ for $t = m/2, ..., T - m/2$. We first show that, under the null hypothesis, $\sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t^*) = \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) + o_{p,\lambda}(1)$. Applying a second order Taylor expansion, we have:

$$\begin{aligned}
& \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t^*) && (35)\\
= \; & \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) \\
& + \sigma^{-1}\frac{1}{2}\left\{ \left[m^{-1}\sum_j \nabla \ln f_j(\widehat{\beta}_t^*) \right] \sqrt{m}\left(\widehat{\beta}_t^* - \beta^*\right) \right. && (36)\\
& \left. - \left[m^{-1}\sum_j \nabla \ln g_j(\widehat{\gamma}_t^*) \right] \sqrt{m}\left(\widehat{\gamma}_t^* - \gamma^*\right) \right\} && (37)\\
& + \sigma^{-1}\left(\widehat{\beta}_t^* - \beta^*\right)' \left[m^{-1}\sum_j \nabla^2 \ln f_j(\ddot{\beta}_t^*) \right] \sqrt{m}\left(\widehat{\beta}_t^* - \beta^*\right) && (38)\\
& - \left(\widehat{\gamma}_t^* - \gamma^*\right)' \left[m^{-1}\sum_j \nabla^2 \ln g_j(\ddot{\gamma}_t^*) \right] \sqrt{m}\left(\widehat{\gamma}_t^* - \gamma^*\right), && (39)
\end{aligned}$$

where $\ddot{\beta}_t^*$ is an intermediate point between $\widehat{\beta}_t^*$ and β^* (and similarly for $\ddot{\gamma}_t^*$). By construction, $m^{-1}\sum_j \nabla \ln f_j(\widehat{\beta}_t^*) = 0$ (and similarly for $\widehat{\gamma}_t^*$); by Assumption FB(2), $\sqrt{T}\left(\widehat{\beta}_t - \beta\right)$ is $O_{p,\lambda}(1)$; therefore, by Assumption FB(5), (36) is $o_{p,\lambda}(1)$, and similarly for (37). Note that

$$\begin{aligned}
& \sigma^{-1}\left(\widehat{\beta}_t^* - \beta^*\right)' \left[m^{-1}\sum_j \nabla^2 \ln f_j(\ddot{\beta}_t^*) \right] \sqrt{m}\left(\widehat{\beta}_t^* - \beta^*\right) \\
= \; & \sigma^{-1}\left(\widehat{\beta}_t^* - \beta^*\right)' \left[m^{-1}\sum_j \nabla^2 \ln f_j(\ddot{\beta}_t^*) - E\left(\nabla^2 \ln f_j(\beta^*)\right) \right] \sqrt{m}\left(\widehat{\beta}_t^* - \beta^*\right) && (40)\\
& + \sigma^{-1}\left(\widehat{\beta}_t^* - \beta^*\right)' E\left(\nabla^2 \ln f_j(\beta^*)\right) \sqrt{m}\left(\widehat{\beta}_t^* - \beta^*\right); && (41)
\end{aligned}$$

by Assumptions FB(2,3), (40) and (41) are both $o_{p,\lambda}(1)$, and similarly for (39). Thus,

$$\sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t^*) = \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta^*) + o_{p,\lambda}(1). \quad (42)$$

Now write

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

By (42) and Assumptions FB(1), FB (4) and FB(5), we have

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

The statement in the proposition then follows from consistency of $\widehat{\sigma}$ for σ. ■

9 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

Table 2. Monte Carlo: Design 1

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

Table 3. Monte Carlo: Design 2

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

Notes to Tables and Figures

Notes to Table 1. The table reports critical values for the fluctuation test in Proposition 3. 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) using $\Pi = [0.15, ..., 0.85]$.

Note to Tables 2-3. The tables report empirical rejection probabilities for the ("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 5 for details.

Notes to Figure 1. The figure refers to the example in Section 2. The solid line in the figure plots the path of time variation of $\Delta KLIC$ arising from β_t varying smoothly while keeping the other parameters constant (time varying parameter case, panel a) and from $\sigma_{x,t}^2$ experiencing a one-time break while keeping the other parameters constant (break in the variance of the regressor case, panel b). Panels c,d show the smoothed ΔKLIC (dotted line). In all panels, the large dot shows the global (or average) ΔKLIC. Panels e,f show the Fluctuation test statistic (solid line) and the boundary lines (dotted lines) in the time varying parameter case and in the break in the variance of the regressor cases, respectively. Panels g,h plot ΔKLIC (solid line) and the one time reversal estimate (dotted line) for the time varying parameter case and in the break in the variance of the regressor cases, respectively.

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

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

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: $\widehat{R}_t = \rho\widehat{R}_{t-1} + (1-\rho)\left\{\overline{\pi}_t + r_\pi\left(\widehat{\pi}_{t-1} - \overline{\pi}_t\right) + r_Y(\widehat{Y}_{t-1} - \widehat{Y}_t^p)\right\} + r_{\Delta\pi}\left(\widehat{\pi}_t - \widehat{\pi}_{t-1}\right) + r_{\Delta Y}((\widehat{Y}_t - \widehat{Y}_t^p) - (\widehat{Y}_{t-1} - \widehat{Y}_{t-1}^p)) + \eta_t^R$, $\overline{\pi}_t = \rho_\pi\overline{\pi}_{t-1} + \eta_t^\pi$. 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 ($\sqrt{var\left(\eta_t^\pi\right)}$).

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 de-trended 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).

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 de-trended 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).

Figure 1(a)

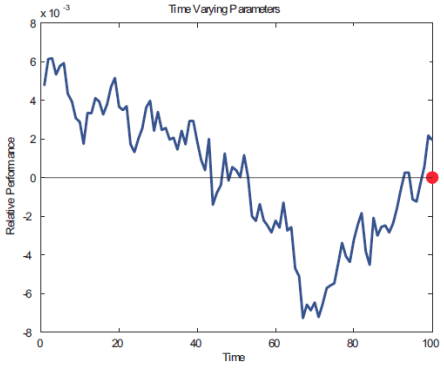

Figure 1(b)

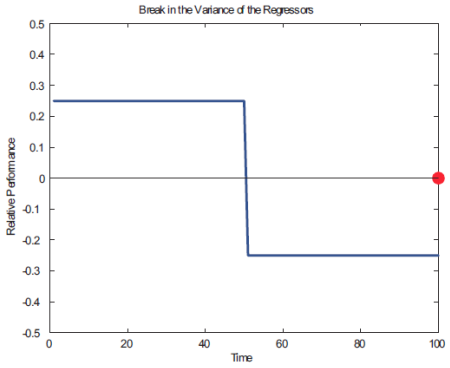

Figure 1(c)

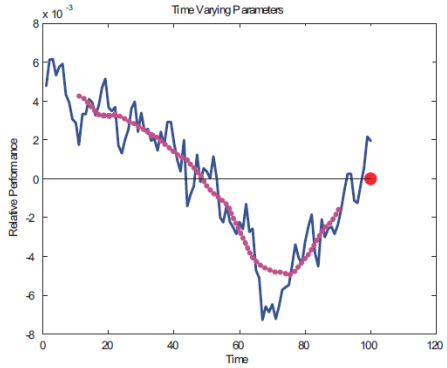

Figure 1(d)

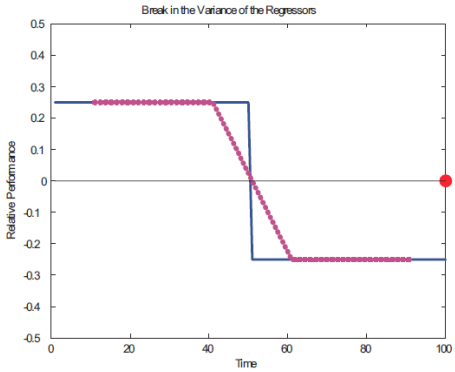

Figure 1(e)

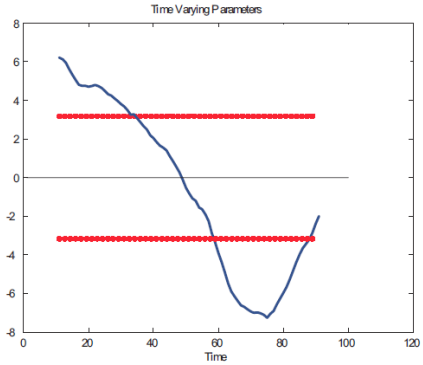

Figure 1(f)

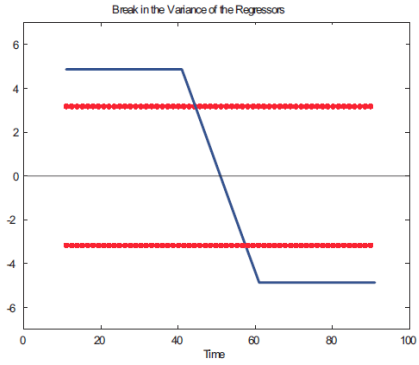

Figure 1(g)

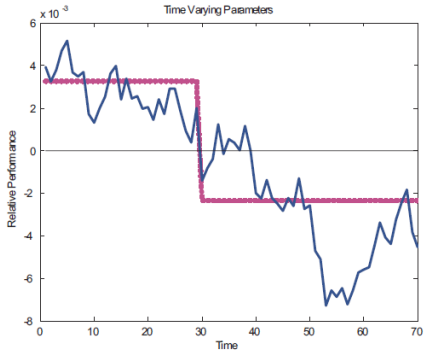

Figure 1(h)

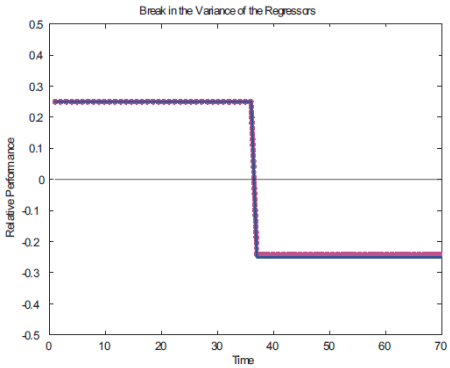

Figure 2(a). Rolling estimates of DSGE parameters (persistence of the shocks).

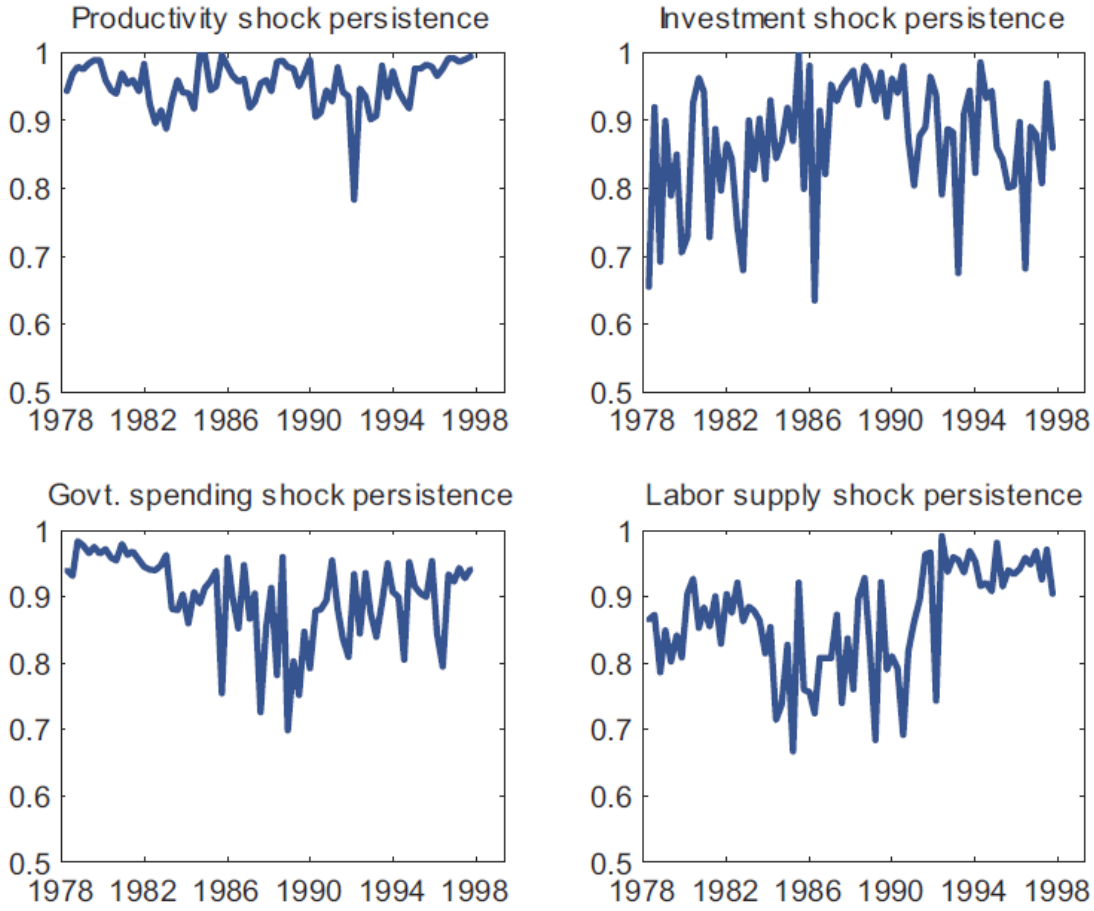

Figure 2(b). Rolling estimates of DSGE parameters (standard deviation of the shocks).

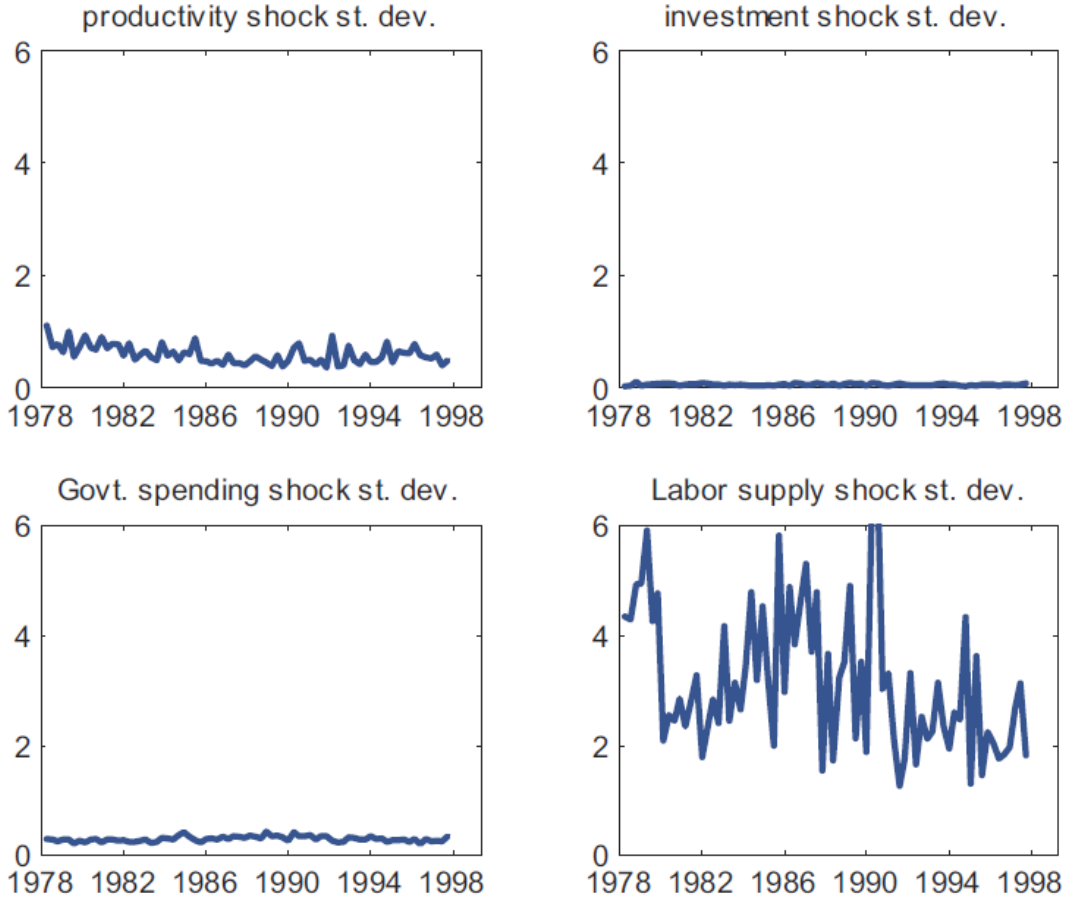

Figure 2(c). Rolling estimates of DSGE parameters (monetary policy parameters).

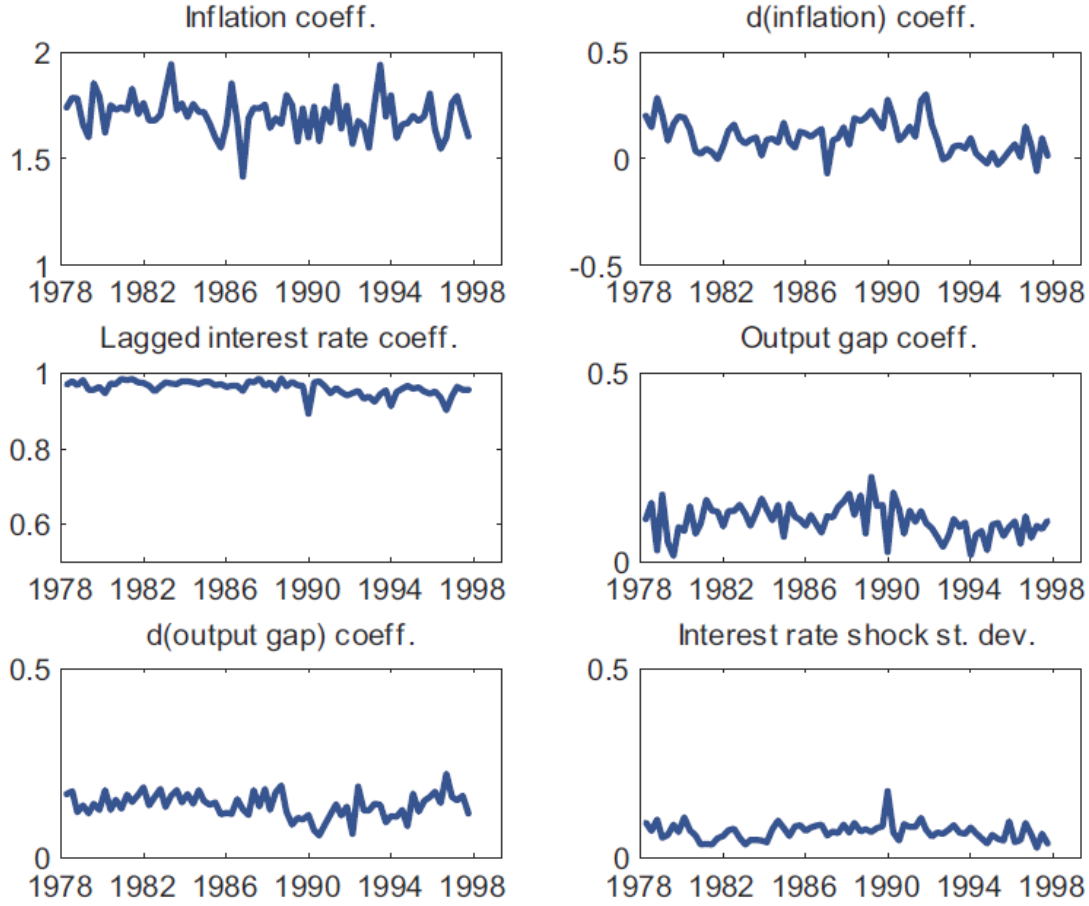

Figure 3. Fluctuation test DSGE vs. BVARs. Full-sample detrending

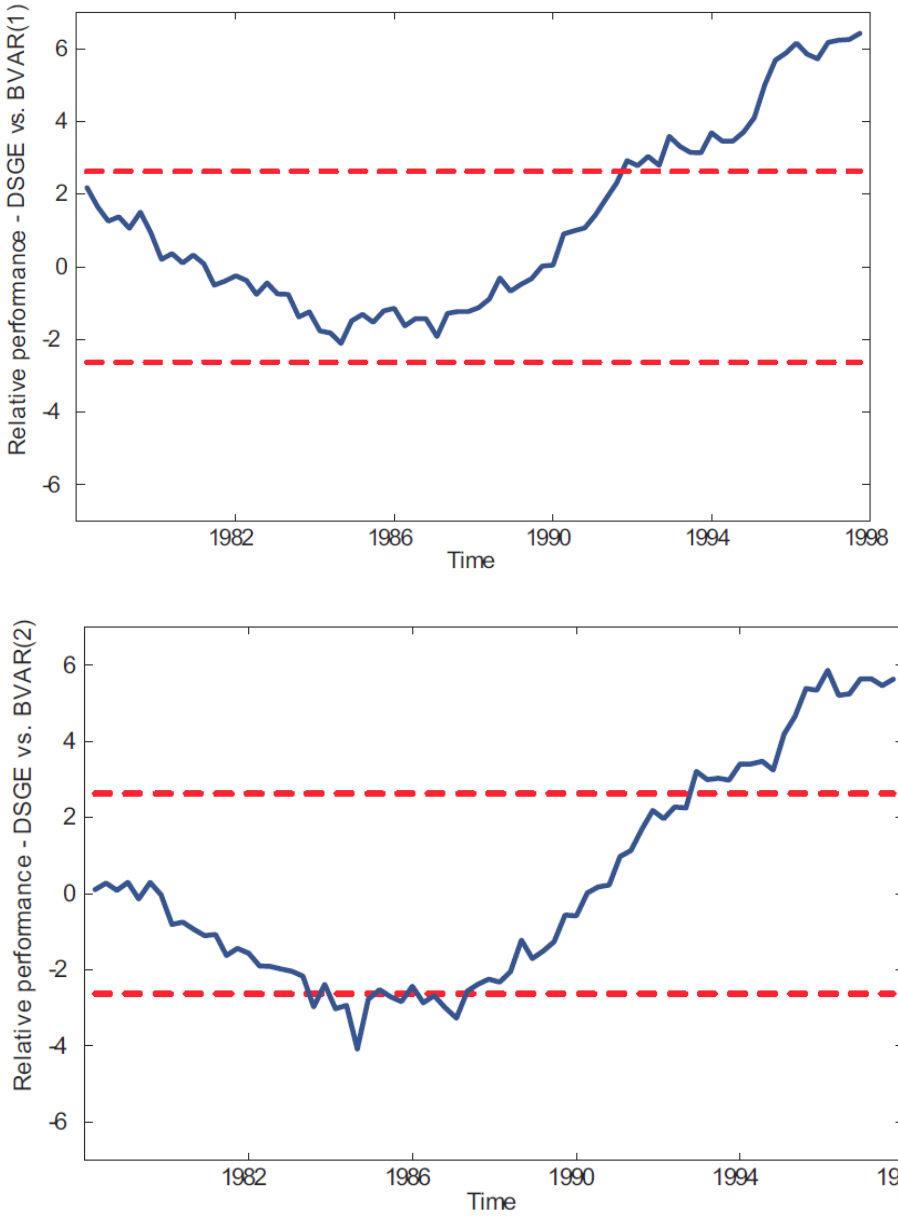

Figure 4. Fluctuation test DSGE vs. BVARs. Rolling sample detrending