Model Comparisons in Unstable Environments

Raffaella Giacomini and Barbara Rossi

UCL and Duke University

May 11, 2010

Abstract

The goal of this paper is to develop formal techniques for analyzing the relative in-sample performance of two competing, misspecified non-nested models in the presence of possible data instability. The central idea of our methodology is to propose a measure of the models' local relative performance: the "local Kullback-Leibler Information Criterion" (KLIC), which measures the relative distance of the two models' (misspecified) likelihoods from the true likelihood at a particular point in time. We discuss estimation and inference about the local relative KLIC; in particular, we propose statistical tests to investigate its stability over time. Compared to previous approaches to model selection, which are based on measures of "global performance", our focus is on the entire time path of the models' relative performance, which may contain useful information that is lost when looking for a globally best model. 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 F. Smets, R. Wouters, W.B. Wu and Z. Zhao for sharing their codes. We also thank 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 and the 2009 NBER-NSF Time Series Conference for useful comments and suggestions. Support by NSF grant 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 investigating instability in the performance of models, rather than 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, and this is not necessarily related to the presence of instability in the model's parameters. 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 estimation and inference about the relative performance of two models over time, and to propose tests that can be used to understand which model performed better at each point in time. Existing econometric tools appear inadequate for answering these questions. On the one hand, model selection tests such as Rivers and Vuong (2002), while allowing the data to have time-varying marginal densities, work under the assumption that there exists a "globally best" model. On the other hand, existing analyses of structural instability solely focus on the parameters of the model, whereas - as the motivating example below will illustrate - it may happen that the relative performance of two models is constant even though their parameters are time-varying or, on the contrary, that the parameters are constant but the relative performance of the models changes over time.

The central idea of our method is to propose a measure of the models' local relative performance: the "local Kullback-Leibler Information Criterion" (KLIC), which measures the relative distance of the two (mis-specified) likelihoods from the true likelihood at a particular point in time. This stands in contrast to the approach of, e.g. Rivers and Vuong (2002), who focus on "global" measures of performance. We then investigate estimation and inference about the local relative KLIC.

Our proposed estimate of the local relative KLIC has a non-parametric flavor. It is obtained by estimating kernel-weighted relative likelihoods, which, importantly, depend on parameters that are also estimated by maximizing kernel-weighted likelihoods. A simple and practically appealing example of such an estimate is obtained by choosing a rectangular kernel, in which case one simply estimates the two models recursively by maximum likelihood (ML) over rolling windows and computes the difference of the average likelihood of each model over the estimation window.

Regarding inference about the local relative performance, we reach several conclusions. First, we show that the dependence of the local performance on unobserved parameters does not affect the asymptotic distribution of the measure of relative performance, as long as the parameters are also estimated locally. Second, in deriving asymptotic inference about our local measure of

performance, we depart from the standard approach in the literature by considering two alternative asymptotic approximations, which we refer to as "fixed bandwidth" and "shrinking bandwidth". We investigate the advantages and limitations of the two approaches and compare the quality of the approximation that they deliver in finite samples. Our Monte Carlo simulations show that the "shrinking bandwidth" approach performs worse than the "fixed bandwidth" approach for sample sizes that are typical for macroeconomists.

In both asymptotic approximations, we show how to estimate the models' relative performance and test for the hypothesis that the two models perform equally well at each point in time. While such procedures are appealing because of their flexibility and ease of implementation, they do not specify an alternative hypothesis, and as a result they may not have optimality properties nor be appropriate for situations in which the time evolution of the relative performance of the models is not smooth. We thus further propose testing and estimation procedures that have optimality properties in the leading and realistic case in which there is a one-time reversal in the relative performance of the models. The Monte Carlo experiment suggests that these tests performs quite well in practice. One important limitation of our approach is that our methods are not applicable when the competing models are nested. This limitation 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.

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) focuses on the different problem of testing models' relative performance for nested and correctly specified models in the presence of instabilities in the parameters. In her approach, the models' relative performance is equal at each point in time only if the parameters that are specific to the larger model are not time varying and equal to zero. Our paper instead focuses on both estimation and testing of the relative performance of non-nested and possibly mis-specified models in unstable environments, which may or may not be related to parameter instabilities. As we will show, the models' relative performance can be stable and equal over time even if the parameters of the competing models change over time. Similarly, the models' relative performance can change over time even if the parameters of the competing models are stable. In a companion paper, Giacomini and Rossi (2010) investigate the related 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. This is also the reason for the introduction of the two different asymptotic approximations with fixed and shrinking bandwidth that we consider in this paper. 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 is also related to, but fundamentally different from, existing tests of parameter stability in that once the measure of local performance is defined, its time variation could in principle be investigated by adapting tools developed in the structural break testing literature to our different context (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). Thus, our tests for the presence of instability in the models' relative performance can be in principle related to this literature, but only after acknowledging that we are testing a different, joint hypothesis that the performance of the models is stable and equal at each point in time. For example, our interest in estimating the path of time variation is similar to Muller and Petalas (2009), but we adopt a different approach, which has a non-parametric flavor. In deriving inference about our proposed estimator, we build on the approach of Wu and Zhao (2007) for the standard "shrinking bandwidth" asymptotic approximation, but we also consider an alternative and novel "fixed-bandwidth" asymptotic approximation which, as it turns out, delivers more reliable inferences for the sample sizes usually available in practice to macroeconomists.

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 notation used throughout the paper, and Section 4 describes the two alternative asymptotic approximations. Section 6 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

Consider an i.i.d. variable y_t with conditional density $h_t : N(\beta_t x_t + \gamma_t z_t, 1)$, $x_t \sim i.i.d. N(0, var(x_t))$, $z_t \sim i.i.d. N(0, var(z_t))$, x_t and z_t are independent and $t = 1, ..., T$. The researcher wants to compare two mis-specified models: model 1, which specifies a density $f_t : N(\beta_t x_t, 1)$ and model 2, with density $g_t : N(\gamma_t z_t, 1)$. 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:

$$\text{Local relative KLIC} : \Delta KLIC_t = E\left[\log h_t/g_t\right] - E\left[\log h_t/f_t\right] = E\left[\log f_t - \log g_t\right], \tag{1}$$

where the expectation is taken with respect to the true density h_t. If $\Delta KLIC_t > 0$, model 1 performs better than model 2 at time t. Simple calculations show that, in our example:

$$\Delta KLIC_t = \frac{1}{2}(\beta_t^2 var(x_t) - \gamma_t^2 var(z_t)) \tag{2}$$

Intuitively, the $\Delta KLIC_t$ 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 $\beta_t^2 var(x_t)$. Similarly, the mis-specification of model 1 is measured by $\gamma_t^2 var(z_t)$. 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 the relative KLIC might be due to the fact that the parameters β_t, γ_t change in ways that affect $\Delta KLIC_t$ differently over time. However, time variation in the local relative KLIC might also occur when the parameters are constant but $var(x_t)$ and $var(z_t)$ change in different ways over time. Finally, note that time variation in the parameters need not necessarily cause time variation in the relative performance; in fact, $\Delta KLIC_t$ can be constant if $\beta_t^2 var(x_t)$ and $\gamma_t^2 var(z_t)$ change in the same way. This may happen when the variances are constant but the parameters change over time in the same way, $\beta_t^2 = \gamma_t^2$, or because changes in the variances offset the relative contribution of the parameter changes to the KLIC.[1]

As two concrete examples, consider the following scenarios of changes in the relative performance of the models. In the first scenario, β_t varies smoothly and γ, $var(x_t)$, $var(z_t)$ are constant, $t = 1, ..., 100$. For example, β_t may evolve according to a random walk. Figure 1(a) shows a possible path for the relative performance. Alternatively, in the second scenario $\beta, \gamma, var(z_t)$ are constant but $var(x_t)$ has a break at $T/2$.

To show why existing approaches to model comparison based on comparing global measures of performance could give misleading conclusions, notice that the test of Rivers and Vuong (2002) would compute the global relative KLIC ($T^{-1} \sum_{t=1}^{T} \Delta KLIC_t$), represented by the large dot in the figures, which compares the average performance of the models over the whole sample. It is clear that in these examples the global relative KLIC is very close to zero, in which case the null hypothesis that the two models perform equally well cannot be rejected. One can see that this occurs because 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. In this paper, we thus advocate focusing on local measures of performance, which will allow the user to recover the full information about the relative performance of the

[1] Note that, for the linear models considered in this example, our null hypothesis is related to requiring that the variances of $\beta_t x_t$ and of $\gamma_t z_t$ are equal at each point in time, which is equivalent to constancy of the sum of squared residuals, an issue examined in Qu and Perron (2007). Section 3 shows that, however, our procedures are more generally applicable to non-linear and general likelihood models.

models over time.

As can be seen from expression (2), the main challenge in estimating the local relative KLIC is its dependence on unknown parameters β_t and γ_t. We solve this problem by estimating both the parameters β_t, γ_t and the measure of relative performance $\Delta KLIC_t$ non-parametrically. When considering inference about the relative performance, we consider two alternative asymptotic approximations. The first is the classical shrinking-bandwidth approximation. 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 therefore consider an alternative approximation where the bandwidth is fixed. In this approximation, consistent estimation of the local relative performance is not possible, but it is nonetheless possible to estimate consistently a "smoothed" version of the local relative KLIC. The object of interest thus becomes:

$$\text{Smoothed local relative KLIC}: \Delta KLIC_t^* = E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\left(\log f_j(\beta_t^*) - \log g_j\left(\gamma_t^*\right)\right)\right], \tag{3}$$

where $K(\cdot)$ is a kernel function, h the bandwidth, and β_t^* and γ_t^* are defined as

$$\begin{aligned}
\beta_t^* &= \arg\max_{\beta} E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\log f_j(\beta)\right], \\
\gamma_t^* &= \arg\max_{\gamma} E\left[\frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{t-j}{Th}\right)\log g_j(\gamma)\right].
\end{aligned}$$

In particular, when using a rectangular kernel the smoothed $\Delta KLIC$ becomes the local average of $\Delta KLIC_t$ over moving windows of size $m = Th$:

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

and β_t^* and γ_t^* are the maximum likelihood "pseudo-true" parameters,

$$\begin{aligned}
\beta_t^* &= \arg\max_{\beta} E\left[\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\log f_j(\beta)\right] \\
\gamma_t^* &= \arg\max_{\gamma} E\left[\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2}\log g_j(\gamma)\right].
\end{aligned}$$

In the example, the local average is $\Delta KLIC_t = \frac{1}{2}\left[\beta_t^{*2}\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2} var(x_j) - \gamma_t^{*2}\frac{1}{m}\sum_{j=t-m/2+1}^{t+m/2} var(z_j)\right]$.

INSERT FIGURE 1 HERE

An alternative measure of the models' relative performance could be obtained by first testing for equal performance over time and, in case of rejection, approximating the relative KLIC assuming a specific form of time variation under the alternative, such as a one-time reversal. In the example previously considered, the dotted lines in Figures 1(c,d) depict $\Delta KLIC_t^*$ for a bandwidth $m/T = 1/5$. The time-path of the relative KLIC under the one-time reversal scenario of Figure 1(d) would be the solid line, which extracts more accurate information about the time variation in relative performance, even though it would provide a poor approximation to the case depicted in Figure 1(c).

Note that, according to the average likelihood ratio test of Rivers and Vuong (2002), represented by the large dot, researchers that are interested in selecting a model for policy purposes or forecast evaluation would be indifferent between the two models. However, the model that fits the data better in the most recent data is model 2, which is the one that should be selected for policy analysis and forecasting. To uncover such changes in the relative performance, we propose a number of statistical tests. In particular, 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. We refer to this test as the Fluctuation test in analogy with the literature on parameter stability testing without assuming an alternative hypothesis (Brown et al. 1975 and Ploberger and Kramer 1992). Figures 1(d,e) 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.

In addition, we also propose tests that are designed to have good power properties against a specific alternative, such as a one-time reversal in the relative performance of the models. Figures 1(g,h) illustrate one of these procedures (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 4.2 for more details.

3 Estimation

In this section, we set forth our approach and propose an estimator of the relative performance of two models over time. We assume that the user has available two possibly mis-specified 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. We measure the relative performance of the two models at each point in time by the local relative KLIC, which represents the relative distance of the two models from the true, unknown, data-generating process at time t :

$$\begin{aligned} \Delta KLIC_t(\theta_t) &= E[\Delta L_t(\theta_t)] = E[\log f_t(\beta_t) - \log g_t(\gamma_t)], \\ \text{for } t &= 1, ..., T, \end{aligned} \tag{4}$$

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

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

where B is a compact parameter space. A similar definition holds for γ_t.

The challenge in estimating the local relative performance of the models is twofold. First, the local relative KLIC is not observable and not necessarily consistently estimable because it is defined as an expectation which could be time-varying. Second, the likelihood difference $\Delta L_t(\theta_t)$ is itself not observable because it depends on θ_t, which cannot in general be estimated consistently, unless one is prepared to make assumptions about the nature of its potential time variation.

We overcome these challenges by considering a non-parametric framework for estimating $\Delta KLIC_t(\theta_t)$. Specifically, we assume

$$\Delta L_t(\theta_t) = \mu_t + \varepsilon_t, \ t = 1, ..., T \tag{5}$$

$$\mu_t = \mu(t/T, \theta(t/T)), \tag{6}$$

where the zero-mean process ε_t is such that a strong invariance principle is satisfied, as discussed in Assumption SB below. We further assume that μ_t and θ_t are generated by smooth functions $\mu(t/T, \theta(t/T))$ and $\theta(t/T)$ defined on $[0, 1]$.

We consider the following nonparametric estimator of μ_t :[3]

$$\widehat{\mu}(\tau) \equiv \widehat{\mu}\left(\tau, \widehat{\theta}(\tau)\right) = \frac{1}{Th} \sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right) \Delta L_t\left(\widehat{\theta}(\tau)\right) \tag{7}$$

[3] Note that the definition is consistent with that in 3 except that we are using the fact that $t = [\tau T]$.

where $K(\cdot)$ is a kernel with $\int K(u)\,du = 1$, h is the bandwidth and $\tau \in [0,1]$ is such that $t = [\tau T]$. We assume that the parameters of the models are also estimated "locally". E.g., the estimator $\widehat{\beta}(\tau)$ for the first model is the solution to

$$\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau - t/T}{h}\right)\nabla \log f_t\left(\widehat{\beta}(\tau)\right) = 0, \tag{8}$$

where $\nabla \log f_t(.)$ denotes the first derivative of the log-likelihood at time t. Note that, in the case of the rectangular kernel considered in Corollary 6 below, this in practice amounts to recursively estimating the parameters of the models by maximum likelihood over rolling windows of length Th.

4 Inference

In this section, we consider the problem of conducting inference about the local relative KLIC. Our goal is to construct confidence intervals and statistical tests of the hypothesis that the models have the same performance at each point in time. Most of this section focuses on the empirically appealing case of a rectangular kernel.

In considering the problem of inference, we depart from the standard approach in the nonparametric literature and consider two alternative asymptotic approximations. One approximation is the traditional "shrinking bandwidth" approximation considered in the literature. In this framework, we will derive asymptotic confidence bands that are simultaneous, and can therefore be used to test the hypothesis of equal performance of the models at each point in time.

The alternative asymptotic approximation, which is new in the literature, considers a fixed bandwidth. In this approximation, our proposed estimator (7) is not consistent for the local relative KLIC, eq. (4), but it consistently estimates a smoothed version of the local relative KLIC, eq. (3), derived as a kernel-weighted average of $\Delta KLIC_t$, which, instead of depending on the local parameter θ_t, depend on the pseudo-true parameters for the chosen kernel. One can thus view this framework as replacing the population object of interest to be the smoothed local relative KLIC instead of the local relative KLIC.

In both approximations, one issue that complicates our analysis is the fact that the likelihood differences depend in a non-linear way on the models' parameters, which are possibly time-varying. As we discuss in more detail below, this fact makes it difficult to obtain valid confidence bands without imposing restrictions on the amount of time variation in other aspects of the data. In our approach, we will impose the assumption that the relative likelihood is globally covariance stationary. One realistic case in which this assumption is satisfied is when the parameters are constant under the null hypothesis; another situation is when parameters change but in ways that ensure that the necessary relative higher moments of the data are constant. When this assumption is not satisfied, one can rely on bootstrap methods such as Cavaliere and Taylor (2005).

4.1 Fixed-bandwidth Asymptotic Approximation: The Fluctuation Test Approach

In this approximation, we focus on the empirically appealing rectangular kernel, which yields the following smoothed local relative KLIC:

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

where $\theta_t^* = (\beta_t^{*\prime}, \gamma_t^{*\prime})'$ and, e.g.,

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

and where $m = Th$.

The smoothed local relative KLIC can be consistently estimated by replacing the expectation with its sample analog and by replacing θ_t^* with the local maximum likelihood estimator (8) computed over rolling windows of size m. Notice that this procedure yields the same estimator of local relative performance $\widehat{\mu}_t$ considered in (7).

Deriving a distribution theory for $\widehat{\mu}$ under this alternative framework, and constructing confidence bands in particular, is not possible if one wants to remain general about how $\Delta KLIC_t^*$ changes over time**.** It is nonetheless possible to construct statistical tests of the hypothesis that $\Delta KLIC_t^*$ equals zero at each point in time. We call this a Fluctuation test.[4]

The Fluctuation test is derived under the following assumptions:

Assumption FB: Let τ be s.t. $t = [\tau T]$ and $\tau \in [0,1]$. (1) $\left\{T^{-1/2} \sum_{j=1}^{[\tau T]} [\Delta L_j(\theta) - E(\Delta L_j(\theta))]\right\}$ obeys a Functional Central Limit Theorem (FCLT) for all $\theta \in \Theta$, Θ compact; (2) there exists a $O(1)$ 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 (and similarly for $\widehat{\gamma}_t$), 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 Strong Uniform Law of Large Numbers, where $\nabla \ln f_j(\beta)$ is a row vector (and similarly for $\nabla \ln g_j(\gamma)$); (4) Under $H_0 : \Delta KLIC_t^* = 0$ for all $t = m/2, ..., T - m/2$, $\sigma^2 = lim_{m\to\infty}$ $E(m^{-1/2} \sum_{j=t-m/2+1}^{t+m/2} \Delta L_j(\theta_t^*))^2 > 0$; (5) $m/T \to h \in (0, \infty)$ as $m \to \infty, T \to \infty$.

Assumption FB(4) imposes global covariance stationarity for the sequence of 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

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

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. Note that this would rule out a situation with $\Delta KLIC_t^*$ being constant in the presence of time-varying parameters and variances. Such situation is easier to accommodate within the shrinking bandwidth approximation, as we discuss in Section 4.3.

The following Proposition provides a justification for the Fluctuation test.

Theorem 1 (Fluctuation Test) *Suppose Assumption FB holds. Consider the test statistic*

$$F_t = \widehat{\sigma}^{-1} m^{-1/2} \sum_{j=t-m/2+1}^{t+m/2} \Delta L_j(\widehat{\theta}_t), \tag{9}$$

$t = m/2, ..., T - m/2$, *where* $\widehat{\sigma}^2$ *is a HAC estimator of* σ^2, *given by, e.g.,*

$$\widehat{\sigma}^2 = \sum_{i=-q(T)+1}^{q(T)-1} (1 - |i/q(T)|) T^{-1} \sum_{j=1}^{T} \Delta L_j \left(\widehat{\theta}_T\right) \Delta L_{j-i} \left(\widehat{\theta}_T\right), \tag{10}$$

$q(T)$ *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* $H_0 : \Delta KLIC_t^* = 0$ *for* $t = m/2, ..., T - m/2$,

$$F_t \Longrightarrow \left[\mathcal{B}(\tau + h/2) - \mathcal{B}(\tau - h/2)\right] / \sqrt{h}, \tag{11}$$

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*

$$\Pr\left\{\sup_{\tau} \left|\left[\mathcal{B}(\tau + h/2) - \mathcal{B}(\tau - h/2)\right] / \sqrt{h}\right| > k_\alpha\right\} = \alpha. \tag{12}$$

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

4.2 Fixed Bandwidth Asymptotic Approximation: Tests Against a One-time Reversal

This section derives tests that are designed for a specific form of time variation in the relative performance of the models, in particular a one-time reversal. Let us define the time path of the relative performance under time variation as follows:

$$E\left(\Delta L_t\left(\beta_t^*, \gamma_t^*\right)\right) = \mu_1^* \cdot 1\left(t \leq [T\delta]\right) + \mu_2^* \cdot 1\left(t > [T\delta]\right)$$

where $t = 1, 2, .., T$, $\theta_t^*(\delta) = \left(\beta_t^{*\prime}, \gamma_t^{*\prime}\right)'$, where δ indexes the time of the reversal as a fraction of the sample size, $\beta_t^* = \beta_1^*(\delta) \cdot 1(t \leq [T\delta]) + \beta_2^*(\delta) \cdot 1(t > [T\delta])$, where
$\beta_1^*(\delta) = \arg\max_\beta E\left[\left(\frac{1}{[T\delta]}\sum_{t=1}^T \ln f_t(\beta) \cdot 1(t \leq [T\delta])\right)\right]$,
$\beta_2^*(\delta) = \arg\max_\beta E\left[\left(\frac{1}{[T(1-\delta)]}\sum_{t=1}^T \ln f_t(\beta) \cdot 1(t > [T\delta])\right)\right]$ (and similarly for $\gamma_t^*(\delta)$, which depends on $g_t(\gamma)$). Let $\beta^*(\delta) = \left(\beta_1^*(\delta)', \beta_2^*(\delta)'\right)'$, and similarly for $\gamma^*(\delta)$, and $\theta^*(\delta) = \left(\beta^*(\delta)', \gamma^*(\delta)'\right)'$.

Consider the problem of testing:

$$H_0 : \mu_1^* = \mu_2^* = 0 \ \forall \delta \in \Pi, \Pi \subset (0,1) \tag{13}$$

versus the alternative

$$H_A : \mu_1^* \neq 0 \text{ or } \mu_2^* \neq 0$$

for some $\delta \in \Pi$, $[\cdot]$ denotes the integer part of $\cdot$.

For simplicity, and in order to derive finite sample optimality results, assume

$$\Delta L_t(\beta_t^*, \gamma_t^*) = \Delta KLIC_t^* + e_t \tag{14}$$

where $\Delta KLIC_t^* \equiv E(\Delta L_t(\beta_t^*, \gamma_t^*))$, $e_t \sim iidN\left(0, \sigma^2\right)$, and denote the log-likelihood of $\Delta L_t(\beta_t^*, \gamma_t^*)$ by $l_t(\mu^*, \delta)$, where $\mu^* \equiv [\mu_1^*, \mu_2^*]'$. In analogy with the parameter instability testing literature, the normality assumption is not a fundamental one here, but is only introduced in order to make statements about the optimality of the test in finite samples.[5] The maximum likelihood estimators of μ_1^* and μ_2^* are

$$\widehat{\mu}_1(\delta) = \frac{1}{[T\delta]}\sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\theta}_t(\delta)\right), \ \widehat{\mu}_2(\delta) = \frac{1}{[T(1-\delta)]}\sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\theta}_t(\delta)\right) \tag{15}$$

and $\widehat{\mu}(\delta) \equiv [\widehat{\mu}_1(\delta), \widehat{\mu}_2(\delta)]'$, $\widehat{\beta}_t(\delta) = \widehat{\beta}_1(\delta) \cdot 1(t \leq [T\delta]) + \widehat{\beta}_2(\delta) \cdot 1(t > [T\delta])$, where $\widehat{\beta}_1(\delta) = \arg\max_\beta \left(\frac{1}{[T\delta]}\sum_{t=1}^T \ln f_t(\beta) \cdot 1(t \leq [T\delta])\right)$, $\widehat{\beta}_2(\delta) = \arg\max_\beta \left(\frac{1}{[T(1-\delta)]}\sum_{t=1}^T \ln f_t(\beta) \cdot 1(t > [T\delta])\right)$, $\widehat{\theta}_1(\delta) \equiv \left[\widehat{\beta}_1(\delta)', \widehat{\gamma}_1(\delta)'\right]'$ (and similarly for $\widehat{\theta}_2(\delta)$), $\widehat{\beta}(\delta) = \left(\widehat{\beta}_1(\delta)', \widehat{\beta}_2(\delta)'\right)'$ (and similarly for $\widehat{\gamma}(\delta)$), $\widehat{\theta}(\delta) \equiv \left[\widehat{\beta}(\delta)', \widehat{\gamma}(\delta)'\right]'$.

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 (13) for the case in which δ is fixed and known. Let $J(\delta)$ be an integrable weight function on the values of δ. The LR statistic for testing the null hypothesis (13), which implies $l_t(0, \delta)$, against a local alternative of the form $l_t\left(\mu^* T^{-1/2}, \delta\right)$ for some $\mu^* \equiv [\mu_1^*, \mu_2^*]'$ is:

$$LR_T = \frac{\int \xi_T\left(\mu^* T^{-1/2}, \delta\right) dQ_\delta(\mu^*) \, dJ(\delta)}{\int \xi_T(0) \, dQ_\delta(\mu^*) \, dJ(\delta)}, \tag{16}$$

[5] Relaxing the normality assumption does not invalidate the test. In particular, these assumptions can be relaxed to allow for general likelihood forms and non iid distributions at the price of increasing notational complexity (involving specifying an additional distribution function for $\Delta L_t(.)$ and notation for its score function) using results such as, e.g., Muller (2010). The main results of this paper would not change by relaxing these assumptions.

where $\xi_T\left(\mu^* T^{-1/2},\delta\right) = N\left(T^{-1/2}\mu_1^* \cdot 1\left(j \leq [T\delta]\right) + T^{-1/2}\mu_2^* \cdot 1\left(j > [T\delta]\right), \sigma^2\right)$ and $\xi_T(0) = N\left(0,\sigma^2\right)$. 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 $\mu^* \in \mathbb{R}^2$, $\delta \in \Pi$.

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

$$\mathcal{I}_{T,\delta} = -\widehat{\sigma}^{-2} \begin{pmatrix} \delta & 0 \\ 0 & (1-\delta) \end{pmatrix}, \tag{17}$$

where we discuss two estimators for $\widehat{\sigma}^2$. The first is an HAC estimator that is valid even in the presence of a structural break in the parameters:

$$\widehat{\sigma}_\delta^2 = \min\left\{\widehat{\sigma}_{1,\delta}^2, \widehat{\sigma}_{2,\delta}^2\right\}, \tag{18}$$

where

$$\begin{aligned}
\widehat{\sigma}_{1,\delta}^2 &= \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|) T^{-1} \sum_{j=1}^{[T\delta]} \left[\Delta L_j\left(\widehat{\theta}_1(\delta)\right) - \widehat{\mu}_1(\tau)\right]\left[\Delta L_{j-i}\left(\widehat{\theta}_1(\delta)\right) - \widehat{\mu}_1(\tau)\right], \\
\widehat{\sigma}_{2,\delta}^2 &= \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|) T^{-1} \sum_{j=[T\delta]+1}^{T} \left[\Delta L_j\left(\widehat{\theta}_2(\delta)\right) - \widehat{\mu}_2(\tau)\right]\left[\Delta L_{j-i}\left(\widehat{\theta}_2(\delta)\right) - \widehat{\mu}_2(\tau)\right],
\end{aligned}$$

and $q(T)$ is a bandwidth that grows with T (e.g., Newey and West, 1987). The second is (10). Note that the former does not require the assumption of global covariance stationarity while the latter does.

Assumptions OT: 1(a) $\Delta L_t(\beta_t^*,\gamma_t^*) = E\left(\Delta L_t(\beta_t^*,\gamma_t^*)\right) + e_t$, *where* $e_t \sim iidN\left(0,\sigma^2\right)$, *and* $\sigma^2 > 0$; *(1b)* $\left\{T^{-1/2}\sum_{j=1}^{[\tau T]} \Delta L_j(\theta)\right\}$ *obeys a Functional Central Limit Theorem (FCLT) for all* $\theta \in \Theta$, Θ *compact; (2) there exists a* $O(1)$ *and positive definite matrix* V_δ *such that* $V_\delta^{-1/2}\sqrt{T}\left(\widehat{\beta}(\delta) - \beta^*(\delta)\right) \xrightarrow{d} N(0,I)$, *as* $T \to \infty$ *uniformly in* δ *(and similarly for* $\widehat{\gamma}(\delta)$*), and* $\beta^*(\delta), \gamma^*(\delta)$ *are interior to the parameter space for every* $\delta \in \Pi$, *where* Π *has closure contained in* $(0,1)$*; (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 similarly for* $\nabla \ln g_t(\gamma)$*); (4) Under* H_0, $E\left(\Delta L_t(\beta_t^*,\gamma_t^*)\right) = 0$ *and the distribution of* $\Delta L_t(\beta_t^*,\gamma_t^*)$ *does not depend on* δ $\forall\theta_t^*$ *and* $\Delta L_t(\beta_t^*,\gamma_t^*)$

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* $c > 0$*; (6)* $sup_{\delta\in\Pi}||\widehat{\mu}(\delta)|| \underset{p}{\rightarrow} 0$ *and* $sup_{\delta\in\Pi}||\widehat{\theta}(\delta) - \theta^*(\delta)|| \underset{p}{\rightarrow} 0$ *under* H_0.[6]

Assumption OT(1a) specifies the structure of the analysis. Assumption OT(1b) 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 4 specifies the null hypothesis. Assumption 5 specifies the weight function over the local alternatives. Assumption 6 assumes that the model is sufficient 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 2 (One-Time Test) *Define the Exponential Wald test,* $ExpW_T^*$*, as:*

$$\begin{aligned} W_T(\delta) &= T\widehat{\mu}(\delta)' H' \left(H\mathcal{I}_{T,\delta}^{-1}H'\right)^{-1} H\widehat{\mu}(\delta) \\ ExpW_T^* &= (1+c)^{-1/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c}W_T(\delta)\right) dJ(\delta) \end{aligned} \tag{19}$$

Under Assumption OT: (i) Under H_0 *described by (13),* $LR_T - ExpW_T^* \underset{p}{\rightarrow} 0$*. (ii) Under the local alternatives in (16), (19) is the test with the greatest weighted average power for the weight functions described in Assumption OT(5).*

The results of Theorem 2 hold in the presence of serial correlation as well as breaks in the variance 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 3 focuses on the limiting case where $c = 0$ and $c = \infty$, and their power properties will be evaluated in Section 5.

Corollary 3 *Suppose Assumption OT holds. Consider the test statistics*

$$\begin{aligned} ExpW_{\infty,T}^* &= \log\frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0} \exp\left(\frac{1}{2}W_T(\delta)\right) d\delta; \\ MeanW_T^* &= \frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0} W_T(\delta)\, d\delta, \text{ where} \end{aligned}$$

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

where $\tau_0 \in \{0.15, ..., 0.85\}$*,* $\widehat{\mu}(\delta)$ *is defined as in (15),* $\mathcal{I}_{T,\delta}$ *is as in (17). Under the null hypothesis* $H_0 : \Delta KLIC_t^* = 0$ *for all* $t = 1, ..., T$*,*

$$ExpW_{\infty,T}^* \Longrightarrow \log\frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0} \exp\left(\frac{1}{2}\frac{\mathcal{BB}(\delta)'\mathcal{BB}(\delta)}{\delta(1-\delta)} + \frac{1}{2}\mathcal{B}(1)'\mathcal{B}(1)\right) d\delta, \tag{20}$$

$$MeanW_T^* \Longrightarrow \frac{1}{1-2\delta_0}\int_{\delta_0}^{1-\delta_0} \left(\frac{1}{2}\frac{\mathcal{BB}(\delta)'\mathcal{BB}(\delta)}{\delta(1-\delta)} + \frac{1}{2}\mathcal{B}(1)'\mathcal{B}(1)\right) d\delta, \tag{21}$$

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

where $t = [\delta T]$ *and* $\mathcal{B}(\cdot)$ *is a standard univariate Brownian motion. 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* $(0.10; 2.44, 4.26)$.

We also provide Sup-type tests for the One-time Reversal in the following proposition:[7]

Proposition 4 (Sup-type Test) *Suppose Assumption OT holds. Let* $QLR^*_T = \sup_{\delta\in\{0.15,\ldots 0.85\}} \Phi_T(\delta)$, $\Phi_T(\delta) = LM_1 + LM_2(\delta)$, *where*

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

$\widehat{\sigma}^2$ *a HAC estimators of the asymptotic variance* $\sigma^2 = var\left(T^{-1} \sum_{t=1}^{T} \Delta L_t(\theta^*_t) \right)$, *for example (10). Consider the null hypothesis*

$$H_0 : \Delta KLIC^*_t = 0,$$

for every $t = 1, 2, ..., T$, *we have:* $QLR^*_T \Longrightarrow \sup_{\delta\in\Pi} \left[\frac{\mathcal{BB}(\delta)'\mathcal{BB}(\delta)}{\delta(1-\delta)} + \mathcal{B}(1)'\mathcal{B}(1) \right]$, *where* $t = [\delta T]$, *and* $\mathcal{B}(\cdot)$ *and* $\mathcal{BB}(\cdot)$ *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:* $(0.05, 9.8257)$, $(0.10, 8.1379)$.

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 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 4 with an overall significance level α:

(i) test the hypothesis of equal performance at each time by using the statistic QLR^*_T from Proposition 4 at α significance level;

(ii) if the null is rejected, compare LM_1 and $\sup_{\delta\in\{[0.15],\ldots[0.85]\}} LM_2(\delta)$, 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 $\sup_\delta LM_2(\delta)$ rejects, then there is evidence that there are instabilities in the relative performance of the two models but neither is constantly better over the

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

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.[8]

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

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

The Fluctuation and the One-time tests have trade-offs. If the researcher is willing to specify the alternative of interest (in this case, a one-time reversal in the relative performance), then the latter test can be implemented and it will have optimality properties. Furthermore, it allows the researcher to estimate the time of the reversal. The Fluctuation test, on the other hand, does not require the researcher to specify an alternative, and therefore might be preferable for researchers who do not have one.[9]

4.3 Shrinking Bandwidth Asymptotic Approximation

In this section we derive results for the local relative KLIC by building on the framework of Wu and Zhao (2007). We make the following assumption:

Assumption SB: (1) $K(\cdot)$ is a symmetric kernel with support $[-w, w]$ which belongs to the class $H(\alpha)$ as in Definition 1 of Wu and Zhao (2007); (2) $\mu \in C^3[0,1]$; (3) The bandwidth h satisfies the condition $Th \to \infty$, $h \to 0$, $\frac{\log(T)^3}{h\sqrt{T}} + Th^7\log(T) \to 0$ and $\frac{\sqrt{Th}}{\ln(T)^3} \to \infty$; (4) $max_{t\leq T}|S_t - \sigma\mathcal{B}(t)| = o_{AS}\left\{T^{1/4}\ln(T)\right\}$, where $S_t = \sum_{i=1}^{t}\varepsilon_i$, $\sigma^2 = \sum_{t=-\infty}^{\infty}E[\varepsilon_0\varepsilon_t]$ and $\mathcal{B}(.)$ is a standard Brownian motion; (5) $\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau-t/T}{h}\right)[\Delta s_t(\theta(\tau)) - E(\Delta s_t(\theta(\tau)))] = O_{as}(1)$ uniformly in $\theta(\tau)$ and τ, where $\Delta s_t(\theta) \equiv \partial\Delta L_t(\theta)/\partial\theta$; (6) $\sigma > 0$; (7) 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;.

Assumptions SB(1)-(4) are similar to those in Wu and Zhao (2007). Assumption SB(4), in particular, deserves further discussion. Even though it is possible to find primitive conditions for

[8] 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^2_1 whereas those for LM_2 are from Andrews (1993).

[9] Note that all the tests could be implemented with a penalty function to penalize overparameterized models in small samples. Typical BIC-type penalty functions would not affect the limiting distribution under the null, and hence our results would be unaffected. One possible advantage of adding the penalty function is that, even when models are nested, our procedures would select the smallest model by construction, independently however of whether the smallest model is correctly specified or not.

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 trivially satisfied under the joint null hypothesis that the models have equal performance and that the parameters are constant. Primitive conditions for Assumption SB(5) can be derived by imposing restrictions on the heterogeneity and dependence of the likelihood scores for the two models. Assumption SB(6) rules out the possibility that the models are nested (see discussion in Rivers and Vuong, 2002).

We can show that the following result holds:

Proposition 5 *Under Assumption SB, asymptotic* $100(1-\alpha)\%$ *simultaneous confidence bands for* μ *are given by*

$$\widehat{\mu}(\tau) - h^2\Psi\widehat{\mu}(\tau)'' \pm \frac{\kappa\,\widehat{\sigma}}{\sqrt{Th}}\left[B_K - \frac{\log\left[\log(1-\alpha)^{-1/2}\right]}{\sqrt{2\log\left(\frac{1}{h}\right)}}\right], \tag{22}$$

$$\Psi = \int K(u)u^2du/2;\ \kappa^2 = \int K^2(u)\,du \tag{23}$$

$$B_K = \sqrt{2\log\left(\frac{1}{h}\right)} + \frac{1}{\sqrt{2\log\left(\frac{1}{h}\right)}}\left[\frac{2-\psi}{2\psi}\log(\log\left(h^{-1}\right)) + \log\left(\frac{C_K^{1/\psi}h_\psi 2^{1/\psi}}{2\sqrt{\pi}}\right)\right], \tag{24}$$

$$C_K = D_K/2\kappa^2;\ D_K = \lim_{\Delta\to 0}\left[|\Delta|^{-\psi}\int\left\{K(x+\Delta)-K(x)\right\}^2dx\right], \tag{25}$$

where

$$\widehat{\mu}(\tau) = \frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau-t/T}{h}\right)\Delta L_t\left(\widetilde{\theta}(\tau)\right), \tag{26}$$

$$\widetilde{\theta}(\tau) = \widehat{\theta}(\tau) - h^2\Psi\widehat{\theta}(\tau)'', \tag{27}$$

$$\widehat{\theta}(\tau) = \left[\widehat{\beta}(\tau)', \widehat{\gamma}(\tau)'\right]', \tag{28}$$

$$0 = \frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau-t/T}{h}\right)\nabla\log f_t\left(\widehat{\beta}(\tau)\right)\ \textit{(and similarly for } \widehat{\gamma}(\tau)\textit{)}, \tag{29}$$

$\widehat{\mu}(\tau)''$ *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 6 *For the rectangular kernel, let $m = Th$ be an even integer. The estimator of the local relative KLIC becomes*

$$\widehat{\mu}(\tau) = \frac{1}{m} \sum_{j=[\tau T]-m/2+1}^{[\tau T]+m/2} \Delta L_j(\widetilde{\theta}(\tau)), \tag{30}$$

$[\tau T] = m/2, ..., T - m/2$, *where $\widetilde{\theta}(\tau)$ is the bias-adjusted maximum likelihood estimator (27) for $\widehat{\theta}(\tau) = \left[\widehat{\beta}(\tau)', \widehat{\gamma}(\tau)'\right]'$ defined by*

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

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

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

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 (30) 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

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

A test of the hypothesis that the models have equal performance at each point in time can be obtained by rejecting the null of equal performance if the horizontal axis is not fully contained within the confidence bands obtained above.

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 three designs for the Data Generating Processes (DGPs). These designs are representative of the features discussed in the main example in Section 1. 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 parameters need not necessarily cause time variation in the relative performance; in fact, $\Delta KLIC_t$ can be constant if the degree

of mis-specification of the competing models changes over time in the same way. Design 2 fits this description. Finally, 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 3.

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

Design 2. $\sigma_{x,t}^2 = \sigma_{z,t}^2 = 1$, $\gamma_t = 1$, $\beta_t = 1 + \beta_A \cdot \left(\widetilde{\beta}_t - T^{-1}\sum_{s=1}^{T}\widetilde{\beta}_s\right)$, $\widetilde{\beta}_t = \beta_A \cdot 0.001 u_t$, $u_t = \rho u_{t-1} + \eta_t$, where $\eta_t \sim N(0,1)$ and $\rho = 0.9$. In this design, one of the parameters (β) changes over time, but in such a way that the expected relative performance of the models is equal over time.

Design 3. $\sigma_{x,t}^2 = 1 + \sigma_A^2 \cdot 1(t > 0.75T)$, $\sigma_{z,t}^2 = 1$, $\beta_t = 1$, $\gamma_t = 1$. 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 to 4 show the empirical rejection frequencies of the various tests for a nominal size of 5%. For the shrinking bandwidth test, we utilize a gaussian kernel with a bandwidth equal to 0.005, which performs very well in design 1 relative to other bandwidths. 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 Shrinking bandwidth 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. Conversely, Table 3 shows that all the tests have fairly good size properties under design 2. Regarding design 3, Table 4 shows that, again, the Shrinking bandwidth 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, 3 AND 4 HERE

Table 5 investigates the robustness of our results for the Fluctuation and One-time Reversal tests in design 1 in the presence of large breaks and when using a HAC covariance estimator implemented

with Andrews' (1991) automatic bandwidth procedure. For the $ExpW^*_{\infty,T}$ and $MeanW^*_T$ tests we show results using either (18) or (10). The table confirms that our procedures are quite robust, and that estimator (18) performs the best.

INSERT TABLE 5 HERE

Finally, Table 6 explore the robustness of our results for the shrinking bandwidth 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 $h = 0.005$.[10]

INSERT TABLE 6 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 atheoretical 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 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 $T = 145$.

[10] 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.

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).[11] 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

We then apply our in-sample 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.

[11] 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.

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).

7 Conclusions

This paper developed estimation and statistical testing procedures for evaluating models' relative performance in unstable environments. Inference and testing are derived in the context of two alternative asymptotic approximations, involving a fixed or a shrinking bandwidth. We also consider optimal tests against a one-time reversal in the models' relative performance. The small sample properties of our procedures are investigated in a series of small Monte Carlo experiments that suggest that the fixed bandwidth approximation is better than the shrinking bandwidth approximation for the sample sizes usually available in practice to macroeconomists. 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. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica* 61, 821–856.

[2] 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.

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

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

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

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

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

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

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

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

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

[15] 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.

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

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

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

[19] 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

Proof of Theorem 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 $\sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t) = \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta_t^*) + o_p(1)$. Applying a mean value expansion, we have:

$$
\begin{aligned}
& \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\widehat{\theta}_t) \qquad (31)\\
= \; & \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta_t^*) \\
& +\sigma^{-1}\frac{1}{2}\left\{\left\{m^{-1}\sum_j \nabla f_j(\ddot{\beta}_t) - E\left[m^{-1}\sum_j \nabla f_j(\ddot{\beta}_t)\right]\right\}\sqrt{m}\left(\widehat{\beta}_t - \beta_t^*\right)\right. \\
& \left. -\left\{m^{-1}\sum_j \nabla g_j(\ddot{\gamma}_t) - E\left[m^{-1}\sum_j \nabla g_j(\ddot{\gamma}_t)\right]\right\}\sqrt{m}\left(\widehat{\gamma}_t - \gamma_t^*\right)\right\} \\
& +\sigma^{-1}\frac{1}{2}\left\{E\left[m^{-1}\sum_j \nabla f_j(\ddot{\beta}_t)\right]\sqrt{m}\left(\widehat{\beta}_t - \beta_t^*\right)\right. \\
& \left. -E\left[m^{-1}\sum_j \nabla g_j(\ddot{\gamma}_t)\right]\sqrt{m}\left(\widehat{\gamma}_t - \gamma_t^*\right)\right\} \\
= \; & \sigma^{-1}m^{-1/2}\sum_j \Delta L_j(\theta_t^*) + o_{as}(1) + o_{as}(1),
\end{aligned}
$$

where $\ddot{\beta}_t$ is an intermediate point between $\widehat{\beta}_t$ and β_t^* (and similarly for $\ddot{\gamma}_t$). By Assumption FB(3) we have that $m^{-1}\sum_j \nabla f_j(\ddot{\beta}_t) - E\left[m^{-1}\sum_j \nabla f_j(\ddot{\beta}_t)\right]$ is $o_p(1)$ and, by Assumption FB(2), $\sqrt{m}\left(\widehat{\beta}_t - \beta_t^*\right)$ is $O_p(1)$ (and similarly for the second model), which proves the first $o_{as}(1)$ in equation (31). The second $o_p(1)$ follows from the fact that $\sqrt{m}\left(\widehat{\beta}_t - \beta_t^*\right)$ is $O_p(1)$ by FB(2) and that $E\left[m^{-1}\sum_j \nabla f_j(\ddot{\beta}_t)\right] - E\left[m^{-1}\sum_j \nabla f_j(\beta_t^*)\right] \underset{p}{\to} 0$ as $m \to \infty$ by Theorem 2.3 of Domowitz and White (1982) given Assumptions FB(2) and FB(3), where $E\left[m^{-1}\sum_j \nabla f_j(\beta_t^*)\right] = 0$ by definition. Now write

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

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

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

where $t = [\rho T]$. The statement in the proposition then follows from the fact that, under H_0, $\widehat{\sigma}$ in (10) is a consistent estimator of σ. ■

Lemma 7 *Let the approximate ML estimators be*

$$\overline{\mu}_1(\delta) \equiv T^{-1/2} \sum_{t=1}^{[T\delta]} \Delta L_t \left(\beta_1^*(\delta), \gamma_1^*(\delta)\right) \tag{32}$$

and

$$\overline{\mu}_2(\delta) \equiv T^{-1/2} \sum_{t=[T\delta]+1}^{T} \Delta L_t \left(\beta_2^*(\delta), \gamma_2^*(\delta)\right). \tag{33}$$

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). By Assumption OT(1a) and (14), for every $\delta \in \Pi$:

$$0 = \frac{1}{[T\delta]} \sum_{t=1}^{[T\delta]} \Delta L_t \left(\widehat{\beta}_1(\delta), \widehat{\gamma}_1(\delta)\right) - \widehat{\mu}_1(\delta) \tag{34}$$

$$0 = \frac{1}{T - [\delta T]} \sum_{t=[T\delta]+1}^{T} \Delta L_t \left(\widehat{\beta}_2(\delta), \widehat{\gamma}_2(\delta)\right) - \widehat{\mu}_2(\delta) \tag{35}$$

>From a mean value expansion of (34):

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

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

where $\ddot{\beta}_1(\delta)$ is an intermediate point between $\widehat{\beta}_1(\delta)$ and $\beta_1^*(\delta)$, and similarly for $\ddot{\gamma}_1(\delta)$. The last two term are $o_p(1)$ by Assumptions OT(3) and OT(4). The first two term in the equality are $o_p(1)$

because $\left(\widehat{\beta}_1(\delta) - \beta_1^*(\delta)\right) T^{1/2}$ is $O_p(1)$ by Assumption OT(4) and $E\left(\frac{\partial}{\partial \beta_1^*(\delta)} \Delta L_t(\theta_1^*(\delta))\right) = 0$ (and similarly for $\frac{\partial}{\partial \gamma_1^*(\delta)} \Delta L_t(.)$). The result follows similarly for (35). ■

Proof of Theorem 2. Define the approximate $ExpW_T$ as:

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

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

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

(b) Using (14), the log-likelihood of $\Delta L_t(\beta_t^*, \gamma_t^*)$ is

$$l_T(\mu_1, \mu_2, \delta) \propto -\frac{1}{2}\sum_{t=1}^{[T\delta]} \frac{\left[\Delta L_t(\beta_t^*, \gamma_t^*) - \mu_1^*\right]^2}{\sigma^2} - \frac{1}{2}\sum_{t=[T\delta]+1}^{T} \frac{\left[\Delta L_t(\beta_t^*, \gamma_t^*) - \mu_2^*\right]^2}{\sigma^2}$$

To derive the ML estimators: from the first order conditions we have, $\forall \delta$:

$$0 = \begin{bmatrix} \frac{\partial}{\partial \mu_1} l_T(\widehat{\mu}_1(\delta), \widehat{\mu}_2(\delta), \delta) \\ \frac{\partial}{\partial \mu_2} l_T(\widehat{\mu}_1(\delta), \widehat{\mu}_2(\delta), \delta) \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{[T\delta]} \left[\Delta L_t\left(\widehat{\theta}_1(\delta)\right) - \widehat{\mu}_1(\delta)\right] \frac{1}{\sigma^2} \\ \sum_{t=[T\delta]+1}^{T} \left[\Delta L_t\left(\widehat{\theta}_2(\delta)\right) - \widehat{\mu}_2(\delta)\right] \frac{1}{\sigma^2} \end{bmatrix}$$

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

$$\widehat{\mu}_1(\delta) = \frac{1}{[T\delta]}\sum_{t=1}^{[T\delta]} \Delta L_t\left(\widehat{\theta}_1(\delta)\right), \ \widehat{\mu}_2(\delta) = \frac{1}{[T(1-\delta)]}\sum_{t=[T\delta]+1}^{T} \Delta L_t\left(\widehat{\theta}_2(\delta)\right). \tag{36}$$

Let approximate estimators be defined as (32) and (33). Lemma 7 shows that the approximate estimators $\overline{\mu}_1(\delta), \overline{\mu}_2(\delta)$ are asymptotically equivalent to $\widehat{\mu}_1(\delta), \widehat{\mu}_2(\delta)$.

Define the approximate LR statistic:

$$\begin{aligned} \overline{LR}_T &= \int \exp\left(\frac{1}{2}\left[H\overline{\mu}(\delta)\right]' \left(H\mathcal{I}_{0,\delta}^{-1} H'\right)^{-1} \left[H\overline{\mu}(\delta)\right]\right) \cdot \\ & \int \exp\left(-\frac{1}{2}\left[H\left(\overline{\mu}(\delta) - \mu^*\right)\right]' \left(H\mathcal{I}_{0,\delta}^{-1} H\right)^{-1} \left[H\left(\overline{\mu}(\delta) - \mu^*\right)\right]\right) dQ_\delta(\mu^*)\, dJ(\delta). \end{aligned} \tag{37}$$

>From (37) and Assumption OT(5), which implies $Q_\delta\left(H\mu^*\right) = N\left(0, cH\mathcal{I}_{0,\delta}^{-1}H'\right)$, we have:

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

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

(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, consider estimation of the variance and, without loss of generality, let $\tau < \delta$. In this case, it is easy to verify that $\widehat{\sigma}_{2,\tau}^2 \underset{p}{\rightarrow} \sigma^2$

$\forall \tau < \delta$. Also,

$$\begin{aligned}
\widehat{\sigma}_{1,\tau}^2 &= \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|)T^{-1} \sum_{j=1}^{[T\tau]} \left[\Delta L_j\left(\widehat{\theta}_1(\tau)\right) - \widehat{\mu}_1(\tau)\right]\left[\Delta L_{j-i}\left(\widehat{\theta}_1(\tau)\right) - \widehat{\mu}_1(\tau)\right] \\
&= \sigma^2 + [\mu_1^* - \widehat{\mu}_1(\tau)] \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|)T^{-1} \sum_{j=1}^{[T\tau]} \left[\Delta L_j\left(\widehat{\theta}_1(\tau)\right) - \mu_1^*\right] \\
&\quad + [\mu_1^* - \widehat{\mu}_1(\tau)] \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|)T^{-1} \sum_{j=1}^{[T\tau]} \left[\Delta L_{j-i}\left(\widehat{\theta}_1(\tau)\right) - \mu_1^*\right] \\
&\quad + \sum_{i=-q(T)+1}^{q(T)-1} (1-|i/q(T)|)T^{-1} \sum_{j=1}^{[T\tau]} [\mu_1^* - \widehat{\mu}_1(\tau)]^2 \\
&\leq \sigma^2 + o_p(1),
\end{aligned}$$

where the $o_p(1)$ follows from the fact that $E\left[T^{-1}\sum_{j=1}^{[T\tau]}\left[\Delta L_{j-i}\left(\widehat{\theta}_1(\tau)\right) - \mu_1^*\right]\right] = 0$. Thus, $\min\left\{\widehat{\sigma}_{1,\tau}^2, \widehat{\sigma}_{2,\tau}^2\right\} \underset{p}{\rightarrow} \sigma^2 \ \forall \tau < \delta$. The result follows from Assumptions OT(1a) and OT(6).
Part (ii) follows from Lemma A4 in Andrews and Ploberger (1994), which guarantees that for the local alternatives in (16), the density $l_t\left(\mu^* T^{-1/2}, \delta\right)$ is contiguous to the density $l_t(0,\delta)$. ■

Proof of Corollary 3. Follows from Theorem 2, where (20), (21) follow from Assumption OT(1b). ■

Proof of Proposition 4. First we show that: (i) $LM_1 = \sigma^{-2}T^{-1}\left[\sum_{t=1}^{T}(\log f_t(\beta_t^*) - \log g_t(\gamma_t^*))\right]^2 + o_p(1)$ and (ii)

$$\begin{aligned}
LM_2(\delta) &= \sigma^{-2}([\delta T]/T)^{-1}(1-[\delta T]/T)^{-1} \\
&\quad [T^{-1/2}\sum_{t=1}^{[\delta T]}(\log f_t(\beta_t^*) - \log g_t(\gamma_t^*)) + \\
&\quad - ([\delta T]/T)\, T^{-1/2}\sum_{t=1}^{T}(\log f_t(\beta_t^*) - \log g_t(\gamma_t^*))]^2 + o_p(1)
\end{aligned}$$

To prove (i), note that

$$\begin{aligned}
LM_1 &= \sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}\left(\log f_t(\widehat{\beta}_1(\delta)) - \log g_t(\widehat{\gamma}_1(\delta))\right) + \\
&\quad + \sigma^{-2}T^{-1}\sum_{t=[T\delta]+1}^{T}\left(\log f_t(\widehat{\beta}_2(\delta)) - \log g_t(\widehat{\gamma}_2(\delta))\right)
\end{aligned} \tag{38}$$

Consider the first summand in (38). By applying a Taylor expansion:

$$
\begin{aligned}
& \sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}\left(\log f_t(\widehat{\beta}_1(\delta)) - \log g_t(\widehat{\gamma}_1(\delta))\right) \\
= \; & \sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}\left(\log f_t(\beta_1^*(\delta)) - \log g_t(\gamma_1^*(\delta))\right) + \\
& +\frac{1}{2}\sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}\left[\nabla \log f_t(\ddot{\beta}_1(\delta)) - E\left(\nabla \log f_t(\ddot{\beta}_1(\delta))\right)\right]\left(\widehat{\beta}_1(\delta) - \beta_1^*(\delta)\right) \\
& -\frac{1}{2}\sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}\left[\nabla \log g_t\left(\ddot{\gamma}_1(\delta)\right) - E\left(\nabla \log g_t\left(\ddot{\gamma}_1(\delta)\right)\right)\right]\left(\widehat{\gamma}_1(\delta) - \gamma_1^*(\delta)\right) \\
& \frac{1}{2}\sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}E\left(\nabla \log f_t(\ddot{\beta}_1(\delta))\right)\left(\widehat{\beta}_1(\delta) - \beta_1^*(\delta)\right) \\
& -\frac{1}{2}\sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}E\left(\nabla \log g_t\left(\ddot{\gamma}_1(\delta)\right)\right)\left(\widehat{\gamma}_1(\delta) - \gamma_1^*(\delta)\right) \\
= \; & \sigma^{-2}T^{-1}\sum_{t=1}^{[T\delta]}\left(\log f_t(\beta_1^*(\delta)) - \log g_t(\gamma_1^*(\delta))\right) + o_p(1) + o_p(1)
\end{aligned}
$$

where $\ddot{\beta}_1(\delta)$ is an intermediate point between $\widehat{\beta}_1(\delta)$ and $\beta_1^*(\delta)$ (similarly for $\ddot{\gamma}_1(\delta)$). Assumption OT(3) ensures that $T^{-1}\sum_{t=1}^{T}\left[\nabla \log f_t(\ddot{\beta}_1(\delta)) - E\left(\nabla \log f_t(\ddot{\beta}_1(\delta)(\delta))\right)\right] \underset{p}{\rightarrow} 0$ (and similarly for component in γ) and, by Assumption OT(2), $T^{-1/2}\left(\widehat{\beta}_1(\delta) - \beta_1^*(\delta)\right) = O(1)$ which proves the first $o_p(1)$. The second $o_p(1)$ is justified by the fact that $E\left(\nabla \log f_t(\ddot{\beta}_1(\delta))\right) - E\left(\nabla \log f_t(\beta_1^*(\delta))\right) \underset{p}{\rightarrow} 0$, where $E\left(\nabla \log f_t(\beta_1^*(\delta))\right) = 0$, and that $T^{-1/2}\left(\widehat{\beta}_1(\delta) - \beta_1^*(\delta)\right) = O(1)$. A similar argument proves that this results holds for the second summand as well as for (ii).

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

$$
\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}\left(\log f_t(\beta_t^*) - \log g_t(\gamma_t^*)\right) \Longrightarrow \mathcal{B}(1) \tag{39}
$$

$$
\begin{aligned}
& \sigma^{-1}\left([\delta T]/T\right)^{-1/2}\left(1 - [\delta T]/T\right)^{-1/2}\left[T^{-1/2}\sum_{t=1}^{[\delta T]}\left(\log f_t(\beta_t^*) - \log g_t(\gamma_t^*)\right)\right. \\
& \left. - \left([\delta T]/T\right)T^{-1/2}\sum_{t=1}^{T}\left(\log f_t(\beta_t^*) - \log g_t(\gamma_t^*)\right)\right] \\
\Longrightarrow \; & \delta^{-1/2}(1-\delta)^{-1/2}\left[\mathcal{B}(\delta) - \delta\mathcal{B}(1)\right] = \delta^{-1/2}(1-\delta)^{-1/2}\mathcal{BB}(\delta)
\end{aligned} \tag{40}
$$

where (39) and (40) are asymptotically independent. Then:

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

and the result follows by the Continuous Mapping Theorem. ■

Proof of Proposition 5. In this proof, let $\widehat{\mu}(\tau)$ in (26) 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 (22) are obtained by showing that for every $u \in \mathbb{R}$ and for $T \to \infty$,

$$\Pr\left\{\frac{\sqrt{Th}}{\sigma}\sup_{\tau\in[wh,1-wh]}\left[\widehat{\mu}\left(\tau,\widetilde{\theta}(\tau)\right) - \mu(\tau,\theta(\tau)) - h^2\Psi\mu''(\tau,\theta(\tau)) - B_K\right] \le \frac{u}{\sqrt{2\ln(1/h)}}\right\} \to \exp(-2\exp(-u)). \tag{41}$$

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

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

To prove this, define

$$\begin{aligned} \widehat{U}_T(\tau) &\equiv \sqrt{Th}\sigma^{-1}\left[\widehat{\mu}\left(\tau,\widetilde{\theta}(\tau)\right) - \mu_t\right] \qquad (42) \\ &= \sigma^{-1}\left\{\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau-t/T}{h}\right)\left[\Delta L_t\left(\widetilde{\theta}(\tau)\right) - \mu_t\right]\right\} \\ &= \sigma^{-1}\left\{\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K\left(\frac{\tau-t/T}{h}\right)\left[\Delta L_t\left(\widetilde{\theta}(\tau)\right) - \Delta L_t(\theta(\tau))\right]\right\} + U_T(\tau) \end{aligned}$$

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:

$$\begin{aligned}
\widehat{\mu}\left(\tau,\widetilde{\theta}(\tau)\right) &\equiv \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta L_t\left(\widetilde{\theta}(\tau)\right) = \qquad (43)\\
&= \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta L_t(\theta(\tau))\\
&\quad + \left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\\
&= \widehat{\mu}(\tau,\theta(\tau)) + \left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)
\end{aligned}$$

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

By substituting (43) in (42), 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

$$\begin{aligned}
&\ln^{1/2}(T)\left|\widehat{U}_T(\tau) - W_T(\tau)\right|\\
\leq\ & \ln^{1/2}(T)\left|U_T(\tau) - W_T(\tau)\right| + \ln^{1/2}(T)\left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right|\\
=\ & o_{as}(1) + \ln^{1/2}(T)\left|\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau-t/T}{h}\right)\Delta s_t\left(\overline{\theta}(\tau)\right)\right]'\left(\widetilde{\theta}(\tau)-\theta(\tau)\right)\right| = o_{as}(1) \qquad (44)
\end{aligned}$$

where the first inequality follows from the triangle inequality, the first $o_{as}(1)$ in the second equality follows from Wu and Zhao (2007, Lemma 2, eq. 29), and the last $o_{as}(1)$ in (44) 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), $E\left[\Delta s_t(\theta(\tau))\right] = 0$ 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. ■

9 Tables and Figures

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

	α	
μ	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 1. 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	Shrinking Bandwidth	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

β_A	Shrinking Bandwidth	Fluctuation	QLR_T^*	Break	$\text{ExpW}_{\infty,T}^*$	MeanW_T^*
0	0.04	0.04	0.05	0.05	0.05	0.05
0.1	0.05	0.04	0.05	0.05	0.05	0.05
0.2	0.05	0.04	0.05	0.05	0.05	0.05
0.3	0.05	0.05	0.05	0.06	0.05	0.05
0.4	0.05	0.04	0.04	0.06	0.05	0.05
0.5	0.05	0.04	0.05	0.05	0.05	0.05
0.6	0.05	0.04	0.05	0.06	0.04	0.04
0.7	0.05	0.05	0.05	0.05	0.05	0.06
0.8	0.05	0.05	0.05	0.06	0.05	0.05
0.9	0.05	0.05	0.05	0.05	0.05	0.05
1	0.05	0.04	0.05	0.06	0.05	0.05

Table 4. Monte Carlo: Design 3

σ_A^2	Shrinking Bandwidth	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 Shrinking-bandwidth ("Shrinking Bandwidth"), 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, Table 3 for design 2 and Table 4 for design 3 – see Section 5 for details.

Table 5. Monte Carlo: HAC Variance Estimation

β_A	Break at $[0.25T]$					Break at $[0.75T]$				
		Estimator (18)		Estimator (10)			Estimator (18)		Estimator (10)	
	Fluct.	$\text{ExpW}^*_{\infty,T}$	MeanW^*_T	$\text{ExpW}^*_{\infty,T}$	MeanW^*_T	Fluct.	$\text{ExpW}^*_{\infty,T}$	MeanW^*_T	$\text{ExpW}^*_{\infty,T}$	MeanW^*_T
0	0.04	0.10	0.07	0.05	0.05	0.04	0.09	0.06	0.05	0.05
0.5	0.78	0.84	0.84	0.44	0.50	0.07	0.41	0.34	0.15	0.15
1	0.99	1	1	0.92	0.94	0.19	0.86	0.80	0.43	0.38
1.5	1	1	1	0.99	0.99	0.35	0.98	0.96	0.67	0.60
2	1	1	1	1	1	0.53	0.99	0.99	0.80	0.75
5	1	1	1	1	1	0.92	1	1	0.97	0.96
10	1	1	1	1	1	0.97	1	1	0.97	0.96
20	1	1	1	1	1	0.98	1	1	0.96	0.95
50	1	1	1	1	1	0.98	1	1	0.94	0.91
100	1	1	1	1	1	0.98	1	1	0.92	0.87

Note to Table 5. The table reports empirical rejection probabilities for the $ExpW^*_{\infty,T}$ and $MeanW^*_T$ tests for design 1 when estimating the variance with either (18) or (10). The table also reports the Fluctuation test ("Fluct.") implemented with estimator (10). All the tests are implemented with a HAC variance estimator using Andrews' (1993) automatic bandwidth selection – see Section 5 for details.

Table 6. 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 6. The table shows empirical rejection probabilities for the Monte Carlo design 1 discussed in Section 5 for the Shrinking Bandwidth case and different bandwidth size (h).

Figure 1(a)

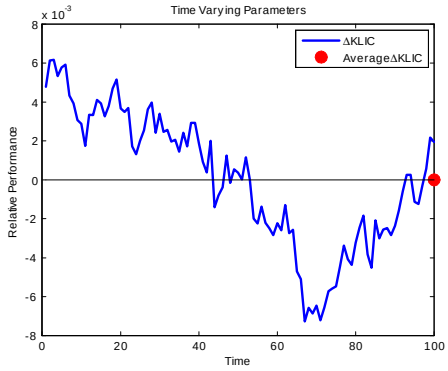


Figure 1(b)

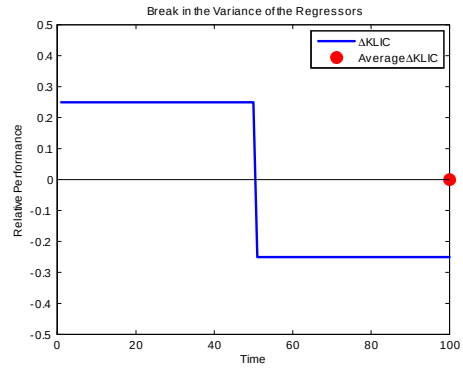


Figure 1(c)

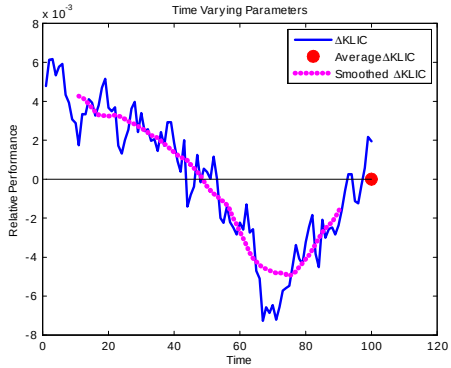


Figure 1(d)

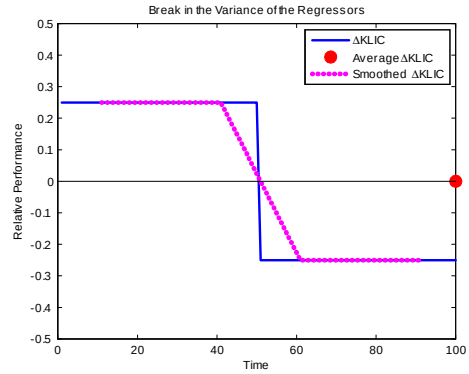


Figure 1(e)

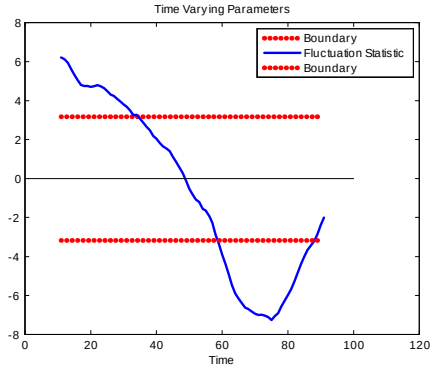


Figure 1(f)

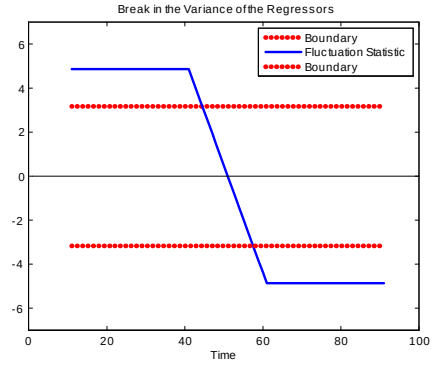


Figure 1(g)

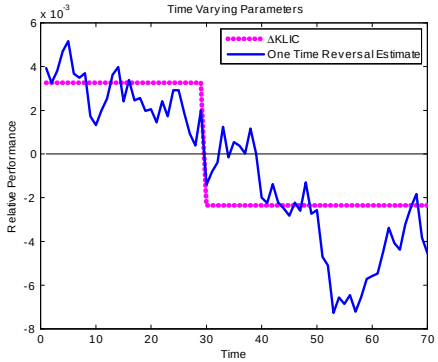


Figure 1(h)

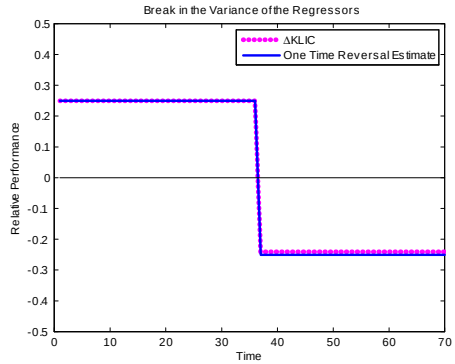
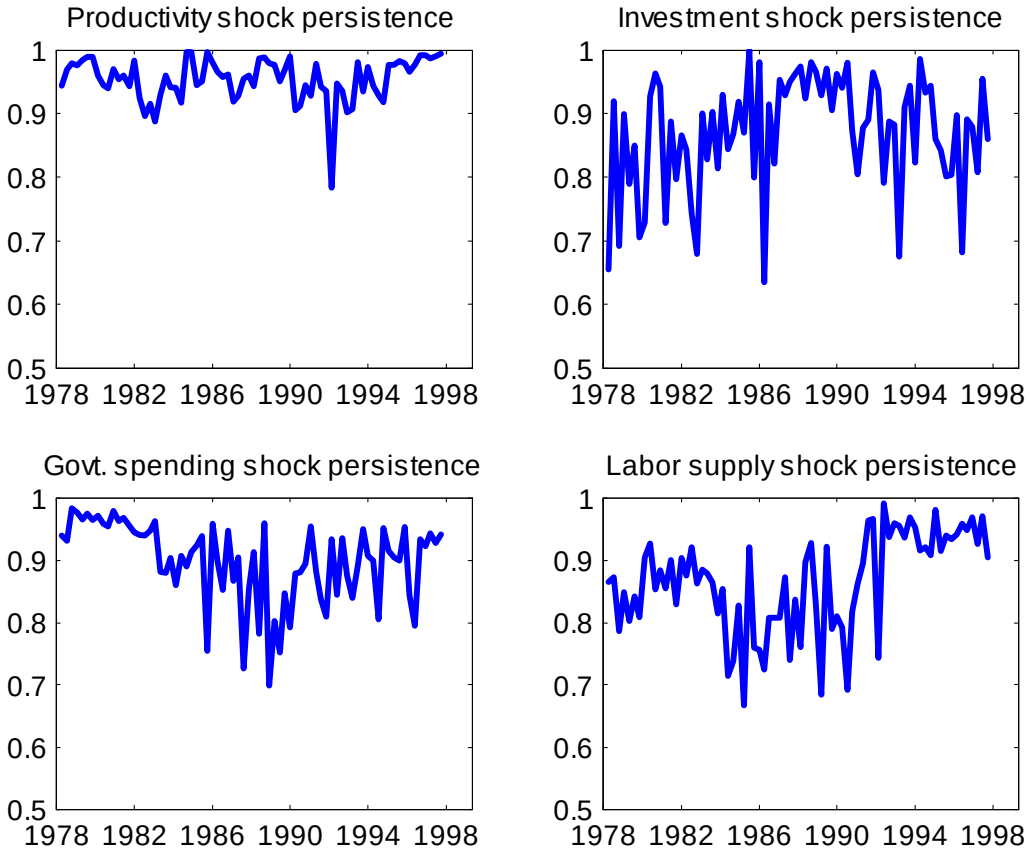
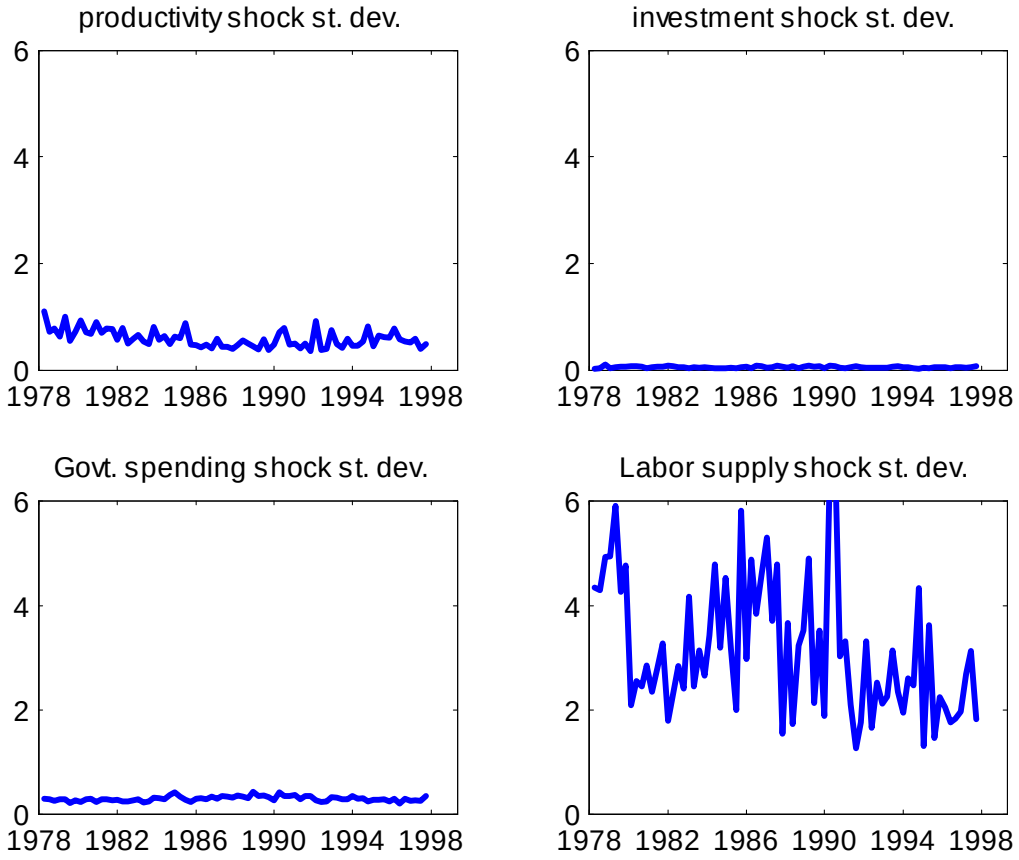


Figure 2(a). Rolling estimates of DSGE parameters (persistence of the shocks).



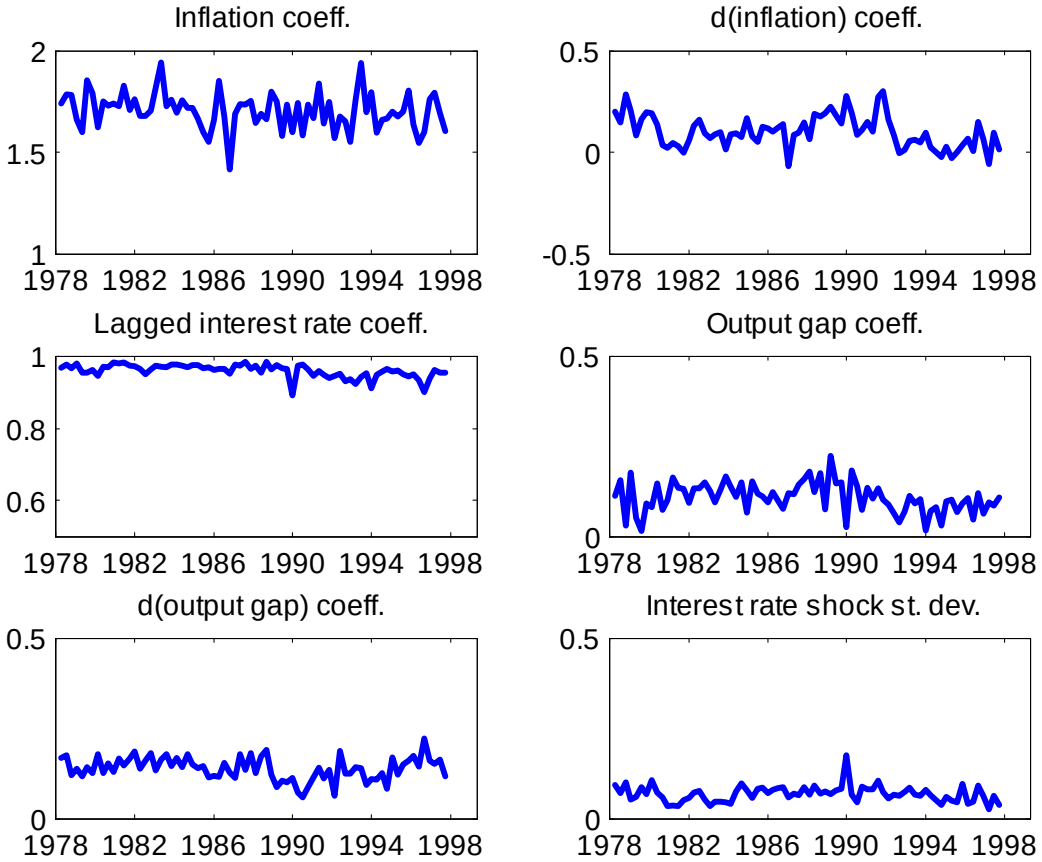
Notes to Figure 2(a). The figure plots rolling estimates of some parameters in Smets and Wouter's (2002) model. See Smets and Wouter's Table 1, p. 1142 for a description.

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



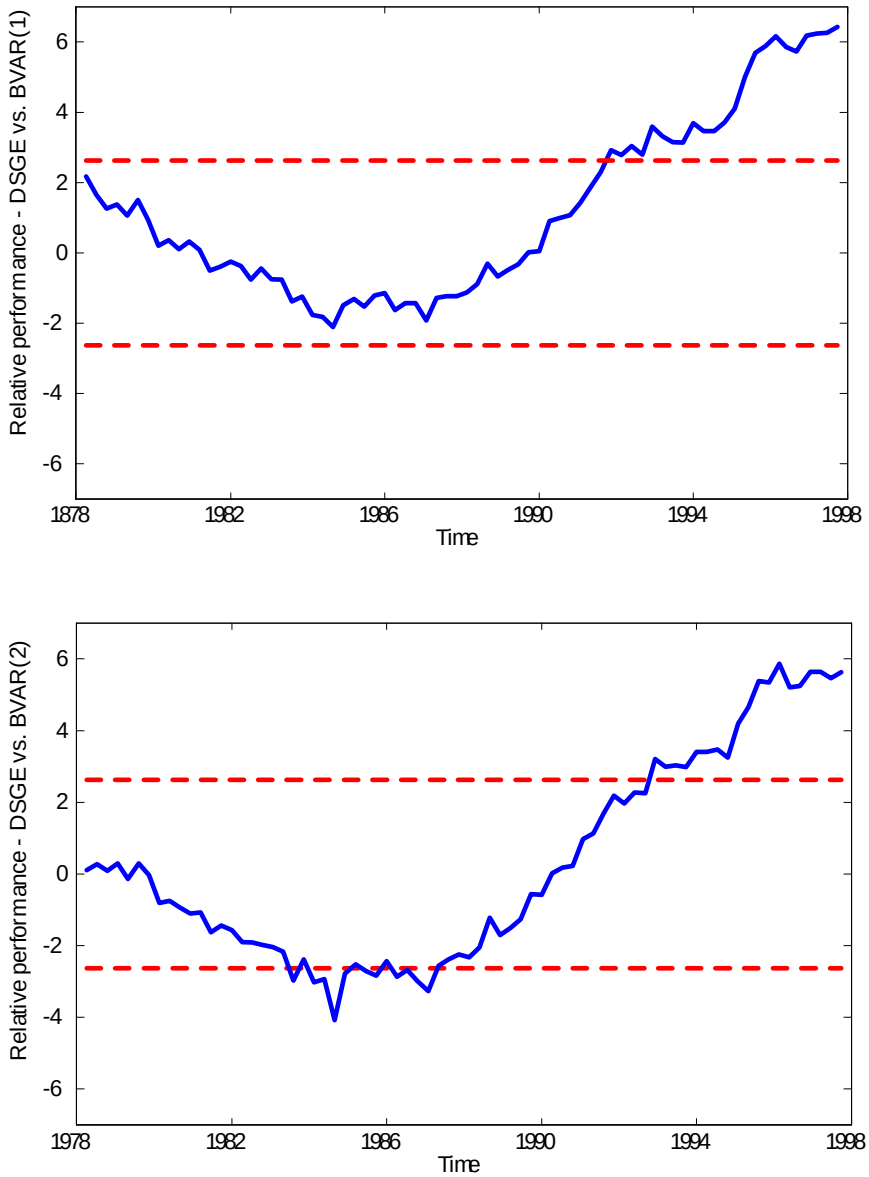
Notes to Figure 2(b). The figure plots rolling estimates of some parameters in Smets and Wouter's (2002) model using full-sample detrended data. See Smets and Wouter's Table 1, p. 1142 for a description.

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



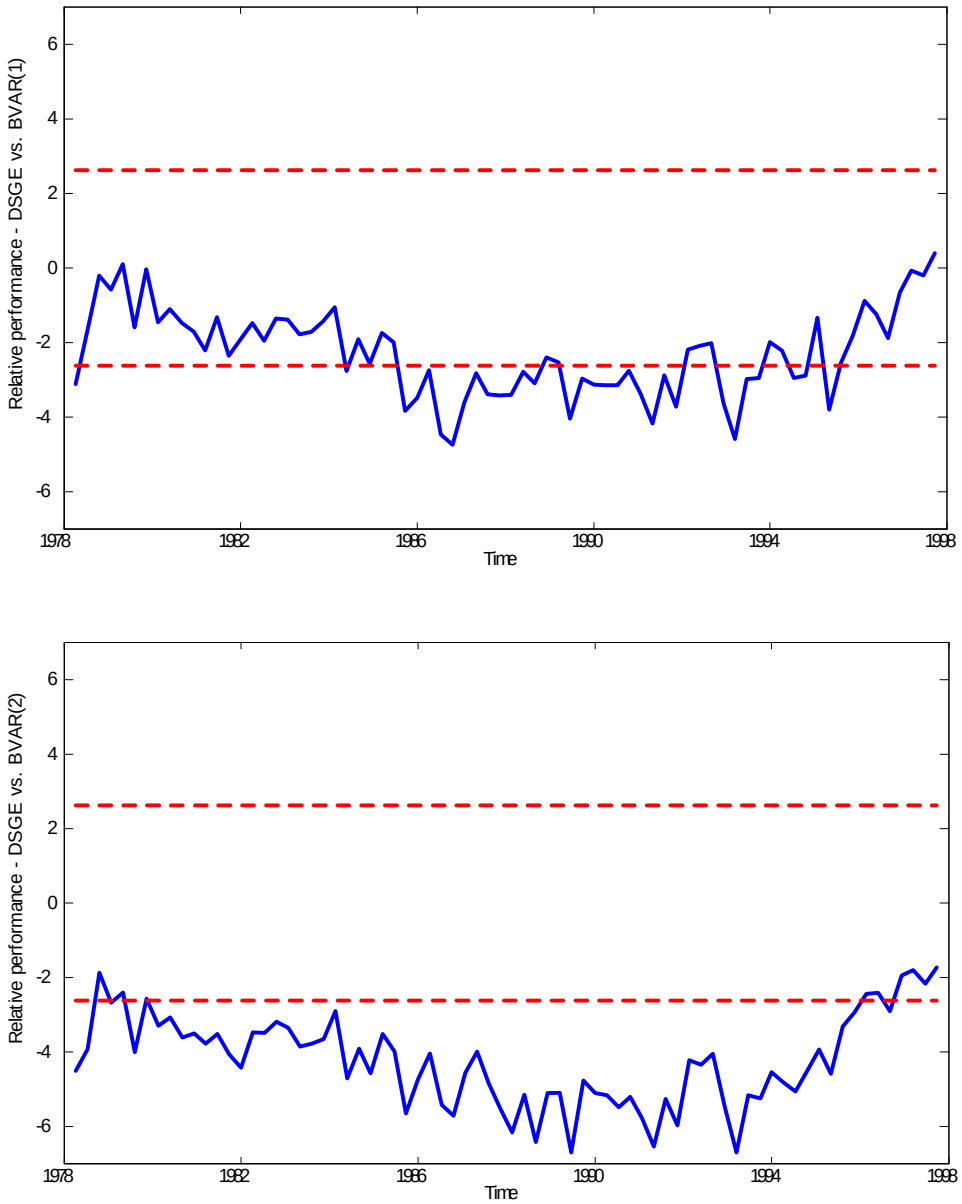
Notes to Figure 2(c). The figure plots rolling estimates of the parameters in the monetary policy reaction function described in Smets and Wouters' (2002) eq. (36), given by: $\widehat{R}_t = \rho \widehat{R}_{t-1} + (1-\rho)\left\{\overline{\pi}_t + r_\pi\left(\widehat{\pi}_{t-1} - \overline{\pi}_t\right) + r_Y\left(\widehat{Y}_{t-1} - \widehat{Y}_t^p\right)\right\} + r_{\Delta\pi}\left(\widehat{\pi}_t - \widehat{\pi}_{t-1}\right) + r_{\Delta Y}\left(\left(\widehat{Y}_t - \widehat{Y}_t^p\right) - \left(\widehat{Y}_{t-1} - \widehat{Y}_{t-1}^p\right)\right) + \eta_t^R$, $\overline{\pi}_t = \rho_\pi \overline{\pi}_{t-1} + \eta_t^\pi$. 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).