Efficient Estimation of the Parameter Path in Unstable Time Series Models

Ulrich K. Müller and Philippe-Emmanuel Petalas

Princeton University

Economics Department

Princeton, NJ, 08544

First Draft: May 2004; This Version: March 2004

Preliminary and Incomplete—Please do not cite or circulate
Comments Welcome

Abstract

The paper investigates asymptotically efficient inference in general time series likelihood models with time varying parameters. Inference procedures for general loss functions are evaluated by a weighted average risk criterion. The weight function focusses on persistent parameter paths of moderate magnitude, and is proportional to the distribution function of a Gaussian random walk. It is shown that asymptotically efficient inference is equivalent to efficient inference in a Gaussian local level model. By implication, estimators of the parameter path and tests of parameter stability are integrated in one unified asymptotic framework. In practice, efficient estimators and test statistics can hence easily be obtained by variants of Kalman smoothing.

JEL Classification: C22, C13, C12, C11

Keywords: Time Varying Parameters, Non-linear Filtering, Non-linear Smoothing, Weighted Average Power, Posterior Approximation, Contiguity

1 Introduction

Instabilities in the parameters of time series models have attracted considerable attention in recent years. The focus of these efforts has mostly been to determine whether instabilities are present: Tests for parameter constancy were proposed and analyzed by Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994), Hansen (2000) and Elliott and Müller (2003), among others. Empirically, instabilities are found to be quite widespread in linear postwar macro-relations, cf. Stock and Watson (1996).

Once instabilities are identified, the natural next step is to document their form. Under the restriction of the instability being of the structural break type, the problem boils down to the determination of the time and magnitude of the break—cf. Bai and Perron (1998), for instance, for a recent contribution. More generally, though, instabilities can arise through all kinds of nonconstant parameter evolution. The question then arises how to estimate the parameter path in unstable time series models.

This paper considers the problem of estimating the parameter path in general, unstable time series likelihood models. We consider estimators that minimize a weighted average risk criterion, where the weighting is over alternative parameter paths. The considered weighting function focusses on persistent parameter paths of the 'Gaussian Random Walk'–type, so that the estimation problem becomes akin to a nonlinear filtering/smoothing problem with a Random Walk state equation.

The concentration on persistent parameter paths is plausible for many applications in economics, as the underlying forces of the parameter changes are often perceived as drifts of preferences, market structures or institutions rather than as sudden shifts. As a modelling strategy, it has found widespread applications: see, for instance, Cooley and Prescott (1976), Harvey (1981), Sims (1993), Stock and Watson (1996), and Primiceri (2004), among many others.

The weight function concentrates on parameter paths of a magnitude for which the data contains only limited information about its form, for all sample sizes. This is the same order of magnitude against which efficient tests of parameter stability have nontrivial power, but not power exactly equal to one. This focus captures the fact that in many applications, it will not be possible to determine the parameter path precisely. In such circumstances, relatively good estimators are those that extract this limited information efficiently.

The main result of this paper is that under some weak regularity assumptions on the likelihood, inference in the unstable general model is asymptotically equivalent to inference in a Gaussian local level model. The score vector of the general model, evaluated at the usual maximum likelihood estimator that ignores the instability, plays the role of the observation equation in the local level model. This in particular implies that for a wide class of loss functions, an asymptotically efficient estimator of the parameter path is obtained by applying a smoother to the sequence of scores. In addition, efficient instability tests that maximize weighted average power over random walk parameter evolutions can be based on this efficient parameter path estimator: An asymptotically efficient test rejects stability when the estimated parameter path is 'too variable' as measured by a specific quadratic form in the estimated parameter instability, just like a standard Wald (1943) statistic.

The rest of the paper is organized as follows. Section 2 gives a heuristic derivation of the asymptotic equivalence between the general filtering/smoothing problem and the local level model, and defines the suggested parameter path estimators and test statistics. Section 3 introduces the estimation problem in detail, discusses the regularity conditions and contains the main results. Section 4 concludes. Proofs are collected in an appendix.

2 Motivation and Definition of Efficient Parameter Path Estimators and Stability Tests

Consider a stable time series model with a log-likelihood function of the form $\sum_{t=1}^{T} l_t(\theta)$, with parameter $\theta \in \Theta \subset \mathbb{R}^k$. Take a true time varying parameter path $\{\theta_0+\delta_t\}_{t=1}^{T}$, where $\{\delta_t\}_{t=1}^{T}$ is normalized by the constraint $\sum \delta_t = 0$, where '$\sum$' denotes a sum over $t = 1, \cdots, T$. The sample information about the path $\{\delta_t\}_{t=1}^{T}$ is then fully contained in the function $\sum l_t(\theta_0+\delta_t)$. Let $\hat{\theta}$ be the maximum likelihood estimator of θ_0 ignoring parameter instability, i.e. $\hat{\theta}$ maximizes $\sum_{t=1}^{T} l_t(\theta)$. Denote the sequence of $k \times 1$ score vectors by $s_t(\theta) = \partial l_t(\theta)/\partial\theta$ and the sequence of $k \times k$ Hessians by $h_t(\theta) = \partial s_t(\theta)/\partial\theta'$. By a second order Taylor expansion

$$\sum l_t(\theta_0 + \delta_t) = \sum [l_t(\hat{\theta} + \delta_t) - s_t(\hat{\theta})'(\hat{\theta} - \theta_0)] - \tfrac{1}{2}(\hat{\theta} - \theta_0)' \left(\sum h_t(\tilde{\theta}_t)\right) (\hat{\theta} - \theta_0)$$

where $\tilde{\theta}_t$ lies on the line segment between $\theta_0 + \delta_t$ and $\hat{\theta} + \delta_t$.

Under standard conditions, $(\hat{\theta} - \theta_0) = O_p(T^{-1/2})$. Suppose the likelihood model is regular enough to ensure a 'Local Law of Large Numbers' for the Hessians, such that for sequences $\{\theta_t\}$ with θ_t close to $\hat{\theta}$ for $t = 1, \cdots, T$, $T^{-1}\sum h_t(\theta_t) + \hat{H} \xrightarrow{p} 0$, where the matrix $\hat{H}$ is defined as $\hat{H} = -T^{-1}\sum h_t(\hat{\theta})$. When the analysis focusses on relatively small parameter variations with $\delta_t = O_p(T^{-1/2})$, the quadratic term can then be approximated by $(\hat{\theta} -$

$\theta_0)' \left(\sum h_t(\tilde{\theta}_t)\right)(\hat{\theta} - \theta_0) \approx -T(\hat{\theta} - \theta_0)'\hat{H}(\hat{\theta} - \theta_0)$, which does not depend on $\{\delta_t\}$. The function $\sum l_t(\hat{\theta} + \delta_t)$ hence contains 'almost' as much information about $\{\delta_t\}$ as the function $\sum l_t(\theta_0 + \delta_t)$.

Now by another Taylor expansion

$$\sum(l_t(\hat{\theta} + \delta_t) - l_t(\hat{\theta})) = \sum s_t(\hat{\theta})'\delta_t + \tfrac{1}{2}\sum \delta_t' h_t(\bar{\theta}_t)\delta_t$$

where $\bar{\theta}_t$ lies on the line segment between $\hat{\theta}$ and $\hat{\theta} + \delta_t$. For parameter variations that are small and persistent, the quadratic term can be approximated by $\sum \delta_t' h_t(\bar{\theta}_t)\delta_t \approx -\sum \delta_t'\hat{H}\delta_t$. The reason is that the persistence in $\{\delta_t\}$ leads to a 'local' averaging of $h_t(\bar{\theta}_t)$: take τ_0 and $\tau_1 > \tau_0$ such that $(\tau_1 - \tau_0)$ is large but $(\tau_1 - \tau_0)/T$ is small, so that δ_t is close to δ_{τ_0} for all $t = \tau_0, \cdots, \tau_1$. Then

$$\sum_{t=\tau_0}^{\tau_1} \delta_t' h_t(\bar{\theta}_t)\delta_t \approx \delta_{\tau_0}' \left(\sum_{t=\tau_0}^{\tau_1} h_t(\bar{\theta}_t)\right)\delta_{\tau_0}.$$

Repeating this local average argument over the whole sample and invoking a Local Law of Large Numbers argument leads to $\sum \delta_t' h_t(\bar{\theta}_t)\delta_t \approx -\sum \delta_t'\hat{H}\delta_t$.

Clearly $\sum s_t(\hat{\theta})'\hat{H}^{-1}s_t(\hat{\theta})$ does not depend on $\{\delta_t\}$. With the approximations in place, we can write

$$\sum(l_t(\hat{\theta} + \delta_t) - l_t(\hat{\theta})) - \tfrac{1}{2}\sum s_t(\hat{\theta})'\hat{H}^{-1}s_t(\hat{\theta}) \approx -\tfrac{1}{2}\sum(\hat{H}^{-1}s_t(\hat{\theta}) - \delta_t)'\hat{H}(\hat{H}^{-1}s_t(\hat{\theta}) - \delta_t) \quad (1)$$

Ignoring constants, the right-hand side of (1) is the the log-likelihood function of the Gaussian random variable $\hat{H}^{-1}s_t(\hat{\theta})$ with mean δ_t and covariance matrix $\hat{H}^{-1}$. The information in the sample about δ_t can therefore be approximately summarized by the observation equation

$$\hat{H}^{-1}s_t(\hat{\theta}) = \delta_t + \nu_t$$

where $\nu_t \sim$i.i.d.$N(0, \hat{H}^{-1})$—a local level observation equation. For a 'state equation' of δ_t that posits its evolution as a *demeaned* Gaussian Random Walk (in order to satisfy the normalization constraint $\sum \delta_t = 0$), the efficient smoothing formula now becomes a variant of the usual Kalman equations. The filtering analogue, of course, can easily be obtained by considering the end-point of the smoothed parameter path of a subset of the data. Also, efficient stability tests that maximize weighted average power against alternatives of Gaussian Random Walk parameter evolution can be based on pseudo likelihood ratio statistics derived from the local level model.

The following section makes these heuristic arguments precise under some fairly general regularity conditions on the likelihood. For many applications, the efficient estimator of the parameter path can be computed as follows:

1. Compute the maximum likelihood estimator of the model, denoted by $\hat{\theta}$, assuming the parameter to be constant through time.

2. For each $t = 1, \cdots, T$, collect the $p \leq k$ elements of the $k \times 1$ score vector $s_t(\hat{\theta})$ that correspond to the time varying parameters in a new $p \times 1$ vector x_t, i.e. $x_t = \Gamma' s_t(\hat{\theta})$ for some $k \times p$ selector matrix Γ that consists of p columns of I_k.

3. Compute the normalized sequence $\tilde{x}_t = \Gamma' \hat{H}^{-1} \Gamma x_t$, $t = 1, \cdots, T$, where $\hat{H} = -T^{-1} \sum_{t=1}^{T} h_t(\hat{\theta})$.

4. Let $z_1 = \tilde{x}_1$, and compute

$$z_t = r_a z_{t-1} + (\tilde{x}_t - \tilde{x}_{t-1}),\ t = 2, \cdots, T$$

where $r_a = 1 - \bar{a}/T$. That is, generate an $p \times 1$ AR(1) process initialized at $\tilde{x}_1$ and innovations $\Delta \tilde{x}_t$.

5. Compute the residuals $\{\tilde{z}_t\}_{t=1}^{T}$ of a linear regression of $\{z_t\}_{t=1}^{T}$ on $\{r_a^{t-1} I_p\}_{t=1}^{T}$.

6. Let $\bar{z}_T = \tilde{z}_T$, and compute

$$\bar{z}_t = r_a \bar{z}_{t+1} + (\tilde{z}_t - \tilde{z}_{t+1}),\ t = 1, \cdots, T-1$$

7. The efficient estimator of the parameter path is now given by $\{\Gamma' \hat{\theta} + \tilde{x}_t - r_a \bar{z}_t\}_{t=1}^{T}$.

This procedure depends on the positive parameter $\bar{a}$, which corresponds to the signal-to-noise ratio in the smoothing problem: The smaller $\bar{a}$, the smoother the estimated parameter path $\{\Gamma' \hat{\theta} + \tilde{x}_t - \bar{z}_t\}_{t=1}^{T}$ becomes. We suggest a default value of $\bar{a} = 10$. This value corresponds roughly to the degree of instabilities empirically found in macro series, cf. Stock and Watson (1998).

In addition, it is the natural choice from a testing perspective: Stability tests that maximize weighted average power over Gaussian Random Walk alternatives achieve power of about 50% against the alternative with a signal-to-noise ratio of that magnitude. An asymptotically efficient test for parameter stability (of the parameters selected by Γ) can be based on the statistic

$$\hat{J} = \sum_{t=1}^{T} (\tilde{x}_t - \bar{z}_t)' \tilde{x}_t$$

where stability is rejected for large values. Asymptotic critical values for $\bar{a} = 10$ are given in Elliott and Müller (2003).

For the computation of the test statistic, it is advisable to replace $\hat{H}$ in step 3 by a consistent estimator of the long-run variance of $\{x_t\}_{t=1}^T$. An attractive choice for such an estimator is derived in Andrews (1991). This replacement yields asymptotically correctly sized tests even when the likelihood model is misspecified, as long as the score represents a valid moment condition.

3 Asymptotically Efficient Inference in Unstable Time Series Models

We begin by introducing some additional notation and definitions. Let $(\mathcal{F}, \mathfrak{F}, P)$ be a probability space, on which all subsequent random elements are defined. The data $y_T = (y_{T,1}, \cdots, y_{T,T})$ in a sample of size T with parameter path $\{\theta + \delta_t\}_{t=1}^T$ is drawn from a parametric model with density

$$f_T(\theta, \delta) = g_T(y_T) \prod_{t=1}^{T} f_{T,t}(\theta + \delta_t),\ \theta + \delta_t \in \Theta$$

with respect to some σ-finite measure μ_T, where θ and δ_t are $k \times 1$, $\delta = (\delta_1', \cdots, \delta_T')' \in \mathbb{R}^{Tk}$ and $g_T(y_T)$ does not depend on (θ, δ). This form of likelihood arises naturally in the 'forecasting error decomposition' of models, where $f_{T,t}(\theta + \delta_t)$ is the conditional likelihood of $y_{T,t}$ given past data, and $g_T(y_T)$ captures the contribution of the evolution of weakly exogenous components (in the sense of Engle, Hendry, and Richard (1983)).

Alternative estimators, or generally decisions, are evaluated via a loss function $L_T : \mathbb{R}^k \times \mathbb{R}^{Tk} \times \mathbb{A}_T \mapsto [0, \bar{L}] \subset \mathbb{R}$, where the action space $\mathbb{A}_T$ is a metric spaces and L_T is assumed Borel-measurable with respect to the product sigma algebra on $\mathbb{R}^k \times \mathbb{R}^{Tk} \times \mathbb{A}_T$. (For reasons that become apparent below, loss is also defined for parameter values outside Θ.) The bound $\bar{L}$ is finite and does not depend on T; this assumption of bounded loss usually has little practical importance, but greatly facilitates the subsequent analysis. When the true parameter evolution is $\{\theta + \delta_t\}_{t=1}^T$ and action $a \in \mathbb{A}_T$ is taken, the incurred loss is $L_T(\theta, \delta, a)$. A typical action could be an estimate of the entire entire parameter path, so that $\mathbb{A}_T = \Theta^T$, or an estimate of the parameter at a specific point in time, in which case $\mathbb{A}_T = \Theta$. Decisions $\hat{a}$ are measurable functions from the data to $\mathbb{A}_T$. The risk of decision $\hat{a}$ given parameter evolution $\{\theta + \delta_t\}_{t=1}^T$ is hence given as $r(\theta, \delta, \hat{a}) = \int L_T(\theta, \delta, \hat{a}) f_T(\theta, \delta) d\mu_T$, which in general depends on δ and θ.

Let Q be a measure on $\mathbb{R}^{Tk}$, and let $w : \Theta \mapsto \mathbb{R}_0^+$ be the Lebesgue density of a random $k \times 1$ vector. For each $\theta \in \Theta$, let $\mathcal{V}_T(\theta) = \{\delta : \delta_t + \theta \in \Theta \forall t\} \subseteq \mathbb{R}^{Tk}$. The Weighted Average Risk of decision $\hat{a}$ is then given by

$$WAR(\hat{a}) = \int_{\Theta} w(\theta) \int_{\mathcal{V}_T(\theta)} r(\theta, \delta, \hat{a}) dQ(\delta) d\theta$$

The weight functions w and Q describe the importance attached to alternative true parameter paths in the overall risk calculations: The weight function w attaches different weights to the average true parameter value $\theta = T^{-1}\sum_{t=1}^{T}(\theta + \delta_t)$, whereas Q describes the focus on deviations from this baseline value.

We make the following condition on w and Q.

Condition 1 *(a) Let $Q = Q_T$ be multivariate normal, such that if $\delta \sim Q_T$ then*

$$\begin{aligned} \delta_t &= \tilde{\delta}_t - T^{-1}\sum_{l=1}^{T}\tilde{\delta}_l, \ \ t = 1, \cdots T \\ \tilde{\delta}_t - \tilde{\delta}_{t-1} &= T^{-1}\Omega^{1/2}\varepsilon_t, \ t = 1, \cdots T, \ \text{and } \tilde{\delta}_0 = 0 \end{aligned}$$

for $\varepsilon_t \sim i.i.d.N(0, I_k)$ for some nonnegative definite, nonrandom $k \times k$ matrix Ω.

(b) w is continuous on Θ.

Under Condition 1, the weighted average risk criterion focusses on parameter paths that are persistent, since under Q, $\{\delta_t\}_{t=1}^{T}$ is distributed as a demeaned Gaussian Random Walk and of relatively small magnitude, as the variance of random walk innovations is given by $T^{-2}\Omega$, for some prespecified Ω.

The concentration on persistent parameter paths is appealing in many applications, as the object of interest are low frequency movements. A structural interpretation of a time-varying regression parameter as a time varying marginal effect, for instance, usually makes more sense if the variation is of a persistent form. Also for forecasting purposes, it is natural to focus on slowly drifting parameters, and then to construct efficient forecasts based on the best guess of the parameter value at the end of the sample.

The sample size dependent choice of the innovation variance $T^{-2}\Omega$ is motivated by a desire to develop procedures that work well when there is relatively little information about the parameter path. For parameter paths of fixed magnitude and persistence, larger samples naturally contain more information, as more adjacent observations can be used to pinpoint the value of the slowly varying parameter at a given date. The sample size dependent choice

$T^{-2}\Omega$ counteracts this effect, making the estimation of the form of the scaled parameter variation $\{T^{1/2}\delta_t\}$ difficult even asymptotically. In this way, the asymptotic arguments derived based on the sequence of weights as described Condition 1 becomes relevant to the small sample problem where there is in fact little information about the parameter evolution. And, of course, parameter variations that are 'small' in the statistical sense of being nontrivial to detect need not to be small in an economic sense. In fact, many instabilities that economists care about, such as those arising from Lucas-critique arguments, the stability of monetary policy or the stability of economic growth have been difficult to determine empirically and are hence 'small' in the statistical sense.

Note that the weight function Q in Condition 1 concentrates its mass in the local neighborhood in which efficient stability tests have nontrivial asymptotic power. A coherent framework of testing for instabilities and a subsequent estimation of its form hence naturally leads to the formulation in Condition 1. Specifically, consider the hypothesis test

$$H_0 : \delta = 0 \quad \text{against} \quad H_1 : \delta \neq 0. \tag{2}$$

Possibly randomized tests φ_T are measurable functions from the data to the interval $[0, 1]$, where $\varphi_T(y_T)$ indicates the probability of rejection when observing y_T. Tests of the same size can then usefully be compared by considering their weighted average power

$$WAP(\varphi_T) = \int_{\mathcal{V}_T(\theta)} \int f_T(\theta, \delta)\varphi_T d\mu_T dQ(\delta) \tag{3}$$

as suggested by Andrews and Ploberger (1994). While (3) is potentially a function of θ, we show below that there exists an efficient test φ_T^* that maximizes (3) for all θ and whose rejection probability under H_0 does not depend on the stable value of θ either.

The nonnegative definite matrix Ω determines the relative variability of the parameters in the model. When a subset of parameters is known to be stable, this knowledge can be incorporated in Ω by setting the appropriate elements equal to zero. In general, the 'larger' Ω, the more weight is put on parameter paths that diverge substantially (in the local neighborhood) from the baseline value θ.

With the weighting of parameter paths specified as the distribution of a (demeaned) Gaussian Random Walk, the problem of finding Weighted Average Risk minimizing actions essentially becomes a nonlinear smoothing exercise. Subject to a measure theoretic qualifier, the loss minimizing decision is the action a that minimizes

$$\frac{\int_\Theta w(\theta) \int_{\mathcal{V}_T(\theta)} f_T(\theta, \delta) L_T(\theta, \delta, a) dQ(\delta) d\theta}{\int_\Theta w(\theta) \int_{\mathcal{V}_T(\theta)} f_T(\theta, \delta) dQ(\delta) d\theta} \tag{4}$$

for each data y_T. With the weighting functions normalized to integrate to unity, this is simply Bayes Rule for minimizing Bayes risk (4), which can be interpreted as finding the action that minimizes the expected posterior loss, i.e. loss integrated with respect to the posterior distributions of (θ, δ) under a prior of (θ, δ) that is proportional to the weights in Condition (1).

A large literature has developed around numerically finding exact posterior distributions in nonlinear filtering/smoothing problems, usually by Monte Carlo simulation techniques. Some of the numerical approximations employ second order Taylor expansions similar to the development in Section 2 above at some stage; see Durbin and Koopman (1997) and Shephard and Pitt (1997), for instance. This paper complements this research by an asymptotic analysis, yielding both a deeper theoretical understanding of the problem and a computationally simple and asymptotically efficient procedure for choosing the risk minimizing action.

For the asymptotic analysis, we require some more notation and regularity conditions on the likelihood $f_T(\theta, \delta)$. We speak of the data being generated 'under θ_0 stable' if the true density of the data is given by $f_T(\theta_0, 0)$, i.e. there is no instability in the parameters, and data generated 'under θ_0 unstable' if the density of the data is proportional to $\int_{\mathcal{V}_T(\theta)} f_T(\theta_0, \delta) dQ(\delta)$. Let $[\cdot]$ indicate the largest lesser integer function, and let '$\xrightarrow{p}$' and '$\Rightarrow$' denote convergence in probability and weak convergence as $T \to \infty$, respectively.

Condition 2 *Under θ_0 stable*

(DIFF) θ_0 is an interior point of Θ, and in some neighborhood $\Theta_0 \subseteq \Theta$ of θ_0, $l_t(\theta)$ is twice differentiable a.s. with respect to θ for $t = 1, \cdots, T$.

(ID) For all $\epsilon > 0$ there exists $K(\epsilon) > 0$ and $\gamma > 0$ such that

$$P(\sup_{||\theta - \theta_0|| \geq \epsilon, \theta \in \Theta} T^{-1} \sum \sup_{||v|| < T^{-1/2+\gamma}, \theta + v \in \Theta} (l_t(\theta + v) - l_t(\theta_0)) < -K(\epsilon)) \to 1$$

(LLLN) Let $\mathcal{B}_T$ be any decreasing neighborhood of θ_0, i.e. $\mathcal{B}_T = \{\theta : ||\theta - \theta_0|| < b_T\}$ for some sequence of real numbers $b_T \to 0$. Then

$$(i) \;\; T^{-1} \sum_{t=1}^{T} \left[\sup_{\theta \in \mathcal{B}_T} h_t(\theta) - \inf_{\theta \in \mathcal{B}_T} h_t(\theta) \right] \xrightarrow{p} 0$$

$$(ii) \;\; \sup_{\lambda \in [0,1]} \left| T^{-1} \sum_{t=1}^{[\lambda T]} [h_t(\theta_0) + H] \right| \xrightarrow{p} 0$$

for some positive definite matrix H.

(FCLT) In the Skorohod metric on the set of cadlag functions on the unit interval

$$T^{-1/2}\sum_{t=1}^{[\cdot T]} s_t(\theta_0) \Rightarrow H^{1/2}W(\cdot)$$

where W is a $k \times 1$ vector Wiener processes.

Condition 2 is a set of fairly standard high level assumptions on the 'forecast error decomposition'-part of the log-likelihood. (DIFF) assumes existence of two derivatives. (ID) is a global identification condition, somewhat strengthened to ensure that even a slightly perturbed evaluation of the likelihood at parameter values far from θ_0 still yields a substantially lower likelihood with high probability. (LLLN) is a Local Law of Large Numbers for the second derivatives h_t. Part (i) controls the average variability of the second derivative h_t as a function of the parameter. It is implied by the more primitive conditions A.2 and A.3 of Andrews (1987). See Gallant and White (1988) and Andrews (1992) for further discussion of these assumptions. Part (ii) assumes linear accumulation of the average information. While quite general, it does typically rule out models with stochastic or deterministic trends. (FCLT) assumes a Functional Central Limit Theorem to hold for the sequence of scores evaluated at the true parameter. As $\{s_t(\theta_0)\}$ constitutes a martingale difference sequence, and typically $T^{-1}\sum_{t=1}^{[\lambda T]} E[s_t(\theta_0)s_t(\theta_0)'] \to \lambda H$ uniformly in $\lambda \in [0,1]$, the high-level assumption (FCLT) can be justified by invoking a FCLT for martingale difference sequences.

Note that Condition 2 puts assumptions only on the stable likelihood model, that is on its behavior when the parameter path is constant. In the presence of time varying parameters, most models generate nonstationary data, to which standard results are not easily applicable. This is especially true for models with weakly exogenous regressors, like Vector Autoregressive Regressions, where parameter instabilities lead to highly complicated interactions between the evolution of the lagged variables and the unstable parameters.

We derive results for the unstable model as an implication of the *contiguity* of the sequence of densities of the unstable model $\{\int_{\mathcal{V}_T(\theta)} f_T(\theta_0,\delta)dQ(\delta)\}_T$ to the sequence of densities $\{f_T(\theta_0,0)\}_T$ of the stable model. Contiguity formalizes the idea of these two densities being close, such that whenever an approximation can be shown to be of order $o_p(1)$ in the stable model, it is also $o_p(1)$ in the unstable model. See Vaart (1998) and Pollard (2001) for a discussion of the concept of contiguity.

To summarize the main results in a compact form, we need to define some additional notation. For $a > 0$, let $\tilde{r}_a = \frac{1}{2}(2 + a^2T^{-2} - T^{-1}\sqrt{4a^2 + a^4T^{-2}}) = 1 - aT^{-1} + o(T^{-1})$. Let A_a

be a $T \times T$ matrix with ones on the main diagonal and $-\tilde{r}_a$ on the first lower subdiagonal, and let $F = A_0^{-1}$. Define the $T \times 1$ vector $e = (1, \cdots, 1)'$, the $T \times T$ matrix $M_e = I_T - ee'/T$, and the $T \times T$ matrix $G_a = M_e - J_a^{-1} + J_a^{-1}e(e'J_a^{-1}e)^{-1}e'J_a^{-1}$, where $J_a = \tilde{r}_a^{-1}FA_aA_a'F'$. Note that if $\delta \sim N(0, a^2T^{-2}M_eFF'M_e)$ and $y_T|\delta \sim N(\delta, M_e)$ (i.e. the model is a scalar demeaned exact local level model with $\Omega = a^2$), then $\delta|y_T \sim N(G_ay_T, G_a)$. [1] Let $P^* \operatorname{diag}(a_1^2, \cdots, a_k^2)P^{*\prime}$ be the spectral decomposition of $\hat{H}^{1/2}\Omega\hat{H}^{1/2}$, and define the $Tk \times Tk$ matrix

$$\Sigma = (I_T \otimes \hat{H}^{-1/2}P^*)\left[\sum_{i=1}^{k} G_{a_i} \otimes (\iota_{k,i}\iota_{k,i}')\right](I_T \otimes P^{*\prime}\hat{H}^{-1/2}) \tag{5}$$

where $\iota_{k,i}$ is the ith column of I_k and '$\otimes$' denotes the Kronecker product. Let Φ_n stand for the distribution $N(0, I_n)$, and define the $Tk \times 1$ vector $\hat{s}$ as $\hat{s} = (s_1(\hat{\theta})', \cdots, s_T(\hat{\theta})')'$.

The main result of the paper is the following Theorem.

Theorem 1 *(a) Assume that the decision $\hat{a}^*$ minimizes*

$$\int\int L_T(\hat{\theta} + T^{-1/2}\hat{H}^{-1/2}u, \Sigma\hat{s} + \Sigma^{1/2}\delta, \hat{a})d\Phi_{Tk}(\delta)d\Phi_k(u)$$

for all y_T. If Conditions 1 and 2 hold for all θ_0 that satisfy $w(\theta_0) > 0$, then for all $\hat{a}$

$$\underline{\lim}_{T\to\infty}[WAR(\hat{a}) - WAR(\hat{a}^*)] \geq 0.$$

(b) Let φ_T^* be the level α test under θ_0 stable, i.e. $\int \varphi_T^* f_T(\theta_0, 0)d\mu_T = \alpha$, that rejects for large values of $\hat{s}'\Sigma\hat{s}$. Then under Conditions 1 and 2, for any other level α test φ_T under θ_0 stable,

$$\underline{\lim}_{T\to\infty}[WAP(\varphi_T^*) - WAP(\varphi_T)] \geq 0.$$

Furthermore, the asymptotic distribution of $\hat{s}'\Sigma\hat{s}$ under θ_0 stable does not depend on θ_0 and is given by Elliott and Müller (2003) in their Lemma 2.

(c) Under Condition 2, with a prior on δ and θ that is proportional to the weights in Condition 1, the total variation distance between the posterior distribution of (θ, δ) and the distribution $N(\hat{\theta}, T^{-1}\hat{H}^{-1}) \times N(\Sigma\hat{s}, \Sigma)$ converges to zero in probability both under θ_0 stable and under θ_0 unstable.

[1] In order to see this, let B_e the $T \times (T-1)$ matrix satifying $B_eB_e' = M_e$ and $B_e'B_e = I_{T-1}$. The unconditional distribution of $B_e'y_T$ then is $N(0, B_e'(a^2T^{-2}FF' + I_T)B_e)$, so that $B_e'\delta|y_T \sim N(B_e'G_aM_ey, (B_e'G_aB_e)^{-1})$, where $B_e'G_aB_e$ satisfies $B_e'G_aB_e = (B_e'(a^2T^{-2}FF' + I_T)B_e)^{-1} + I_{T-1}$. The statement now follows from Lemma 4 of Elliott and Müller (2003).

Comments: 1. The matrix Σ can be understood as describing k local-level models in the elements of the normalized score sequence $\{P^{*\prime}\hat{H}^{-1/2}s_t(\hat{\theta})\}_{t=1}^{T}$, with signal-to-noise ratios given by $a_1, \cdots, a_k$, respectively. The optimal smoother for the true path of $\{\delta_t\}_{t=1}^{T}$ is given by $\Sigma\hat{s}$, with an uncertainty described by a $Tk \times 1$ mean zero multivariate normal with covariance matrix Σ. The three parts of Theorem 1 are different implications of the asymptotic equivalence between the general model and these k scalar Gaussian local level models. In addition, the asymptotic uncertainty about the average level of the parameter path θ is described by a multivariate normal with mean $\hat{\theta}$ and covariance matrix $T^{-1}\hat{H}^{-1}$, just as in stable likelihood models. Note that this uncertainty concerning θ is independent of the choice of weight function w.

2. Part (a) establishes that for arbitrary bounded loss functions, the decision that minimizes risk in the approximate Gaussian local level model is also asymptotically optimal in the true model. Note that loss may be defined arbitrarily (subject to the bounding condition) for parameter values outside Θ, allowing the local level problem to be made entirely spherical. For the wide range of loss functions for which one would choose the smoothed path in a Gaussian local level model, an asymptotically efficient estimator is hence given by $\hat{\theta} \otimes e + \Sigma\hat{s}$. Note that such loss functions include those that consider a weighted average of symmetric losses incurred by estimation errors in the parameter value for all $t = 1, \cdots, T$, but also a symmetric loss function that focusses entirely on, say, the value of the parameter at date T. This kind of loss function might arise naturally in a forecasting problem.

For more general losses and decision problems, the asymptotically efficient decision can still be obtained by implementing the efficient decision in the approximate local level model. This typically represents a dramatic computational simplification.

3. Part (b) spells out the implications of the approximation for efficient tests of the null hypothesis of parameter stability (2). With the smoother of the parameter instability δ being $\Sigma\hat{s}$, and its uncertainty described by a zero mean multivariate normal with covariance matrix Σ, the efficient test statistic is simply of the usual Wald form $(\Sigma\hat{s})'\Sigma^{+}(\Sigma\hat{s}) = \hat{s}'\Sigma\hat{s}$, where Σ^{+} denotes the Moore-Penrose inverse. Efficient estimation and testing in (potentially) unstable models is hence unified in one coherent framework: Efficient instability tests are based on a quadratic form in the efficient estimator of the instability.

In a linear Gaussian time series model, the test statistic $\hat{s}'\Sigma\hat{s}$ becomes the statistic $\hat{J}$ derived by Elliott and Müller (2003). For that special case of a Condition 2 model, they prove asymptotically efficiency for a wider class of weighting functions on δ, that includes Q of Condition 1.

4. Part (c) describes the approximation result in Bayesian terms. The result for the posterior distribution of θ is standard in stable models—see, for instance, Schervish (1995). The approximation result for the posterior of δ might be more surprising: despite the fact that the dimension of δ is the same as that of the data, the posterior of δ becomes arbitrarily close to the Tk dimensional multivariate normal distribution $N(\Sigma\hat{s}, \Sigma)$. This is a much stronger statement than a convergence in distribution of, say, the posterior of $T^{1/2}\delta_{[\cdot T]}$ viewed as an element of the space of cadlag functions on the unit interval.

In practice, part (c) is useful for Bayesian analyses as it provides a simple to compute approximation to the posterior of the unstable parameter path. Even if the exact small sample posterior is required, the approximation of Theorem 1 can still be helpful, as numerical methods typically require a reasonable initial guess of the posterior distribution. In the appendix, we provide a simple algorithm for generating random variables with distribution $N(\Sigma\hat{s}, \Sigma)$.

When applying part (a) and (b) of Theorem 1, the question arises how to choose Ω. An attractive default choice is $\Omega = \mathrm{diag}(\bar{a}^2\Gamma' H^{-1}\Gamma, 0_{k-p})$ for some $\bar{a} > 0$, where Γ is the $k \times p$ selector matrix that selects the $p \leq k$ (potentially) time varying elements of θ introduced in Section 2. This choice of Ω equates the degree of uncertainty about the unstable elements of δ in any given direction (in $\mathbb{R}^p$) with the average sample information about that direction, as under Condition 2, H^{-1} is the information matrix of θ. It hence leads to equal signal-to-noise ratios in all unstable directions. It is also the only choice that yields asymptotic results that do not depend on a particular parametrization. Nyblom (1989), Andrews and Ploberger (1994) and Elliott and Müller (2003) argue for the same choice for their efficient testing procedures.

Typically, of course, H is unknown, but it can be consistently estimated under θ_0 stable. As long as the loss function L_T does not focus excessively on the scale of the estimated parameter instability,[2] a replacement of Ω by $\hat{\Omega}$ in the definition (5) of Σ, yielding $\hat{\Sigma}$, does not affect the optimality results.

Theorem 2 *Suppose $\hat{\Omega} \xrightarrow{p} \Omega$, and $\tilde{H} \xrightarrow{p} H$ under θ_0 stable. Then part (b) of Theorem 1 holds with Σ replaced with $\hat{\Sigma}$, and if $\sup_{\theta,\delta,a} |L_T(\theta, (I_T \otimes (I_k + \Delta_T))\delta, a) - L_T(\theta, \delta, a)| \to 0$ for all sequences of $k \times k$ matrices $\Delta_T \to 0$, then also part (a) of Theorem 1 holds with Σ replaced by $\hat{\Sigma}$ and $\hat{H}$ replaced by $\tilde{H}$.*

[2]This qualifier is necessary, as one could construct pathological loss functions that 'detect' the small difference in scale between $\delta \sim Q$ and $(I_T \otimes \hat{\Omega}^{1/2}(\Omega^{+})^{1/2})\delta \sim Q$.

4 Conclusions

Most economic relationships are potentially unstable over time. In empirical work, this translates into time varying parameters of estimated models. It has long been recognized (cf. Cooley and Prescott (1976)) that it would often be desirable to keep track of this potential instability. Going beyond time variation in the coefficients of Gaussian linear regression models, however, typically leads to major numerical and computational complications.

This paper extends the applicability of straightforward Kalman formulae to much more general models while preserving asymptotic efficiency. As long as the parameter evolution is thought of as a Gaussian Random Walk of moderate magnitude, a local level model with the score as the observation is an asymptotically efficient summary of the sample information. This implies that asymptotically efficient estimators of the parameter path, as well as efficient tests of parameter stability, can easily be obtained from the sequence of scores.

The results of this paper are hence not only of theoretical interest, but they arguably add a useful tool to the applied econometrician's toolbox. At least for a 'first look' at potentially unstable time series models, the procedures suggested here constitute an attractive alternative to numerical approximations to the exact solution: they are computationally straightforward, they have rigorous asymptotic justifications, and they embed efficient tests of parameter stability and efficient parameter path estimators in one coherent framework.

5 Appendix

Proof of Theorem 1:

Proof of part (c). For $\theta_0 \in \Theta$ such that $w(\theta_0) > 0$, let $LR(\theta, \delta) = f_T(\theta, \delta)/f_T(\theta_0, 0)$. The posterior density of (θ, δ) given the data can be written as

$$\frac{\Theta \mathcal{V}_T(\theta) w(\theta) LR(\theta, \delta)\, dQ(\delta) d\theta}{\int_\Theta w(\theta) \left[\int_{\mathcal{V}_T(\theta)} LR(\theta, \delta)\, dQ(\delta)\right] d\theta}.$$

Let $u(\theta) = T^{1/2}(\theta - \theta_0)$, $\hat{u} = T^{1/2}(\hat{\theta} - \theta_0)$ and $\widehat{LR}_T(\theta, \delta) = \exp[\sum s_t(\hat{\theta})'\delta_t - \frac{1}{2}\sum \delta_t' \hat{H} \delta_t + \hat{u}'\hat{H}u(\theta) - \frac{1}{2}u(\theta)'\hat{H}u(\theta)]$. Then the density of $N(\hat{\theta}, T^{-1}\hat{H}^{-1}) \times N(\Sigma\hat{s}, \Sigma^{+})$ can similarly be written as

$$\frac{w(\theta_0)\, \widehat{LR}_T(\theta, \delta) dQ(\delta) d\theta}{w(\theta_0) \int\int \widehat{LR}_T(\theta, \delta) dQ(\delta) d\theta}$$

(see Lemma (8)). The total variation distance between the posterior distributions converges in probability to zero if the L_1 distance between their densities converges, i.e. if

$$\int\int \left| \frac{\Theta \mathcal{V}_T(\theta) w(\theta) LR_T(\theta, \delta)}{\int_\Theta \int_{\mathcal{V}_T(\theta)} w(\theta) LR_T(\theta, \delta) dQ(\delta) d\theta} - \frac{w(\theta_0)\, \widehat{LR}_T(\theta, \delta)}{w(\theta_0) \int\int \widehat{LR}_T(\theta, \delta) dQ(\delta) d\theta} \right| dQ(\delta) d\theta \xrightarrow{p} 0.$$

For any loss function L, let

$$R_T := \int\int \left| \Theta_u \mathcal{V}_T(u) w(\theta_0 + T^{-1/2}u) \frac{LR_T(u, \delta)}{D} - w(\theta_0) \frac{\widehat{LR}_T(u, \delta)}{\widehat{D}} \right| L(\theta_0 + T^{-1/2}u, \delta, a) dQ(\delta) du$$

where $\Theta_u = \{u : \theta_0 + T^{-1/2}u \in \Theta\}$, $D = \int_{\Theta_u} \int_{\mathcal{V}_T(u)} w(\theta_0 + T^{-1/2}u) LR_T(u, \delta) dQ(\delta) du$, and $\widehat{D} = w(\theta_0) \int\int \widehat{LR}_T(u, \delta) dQ(\delta) du$, and notice that the L_1 distance between the posterior densities is R_T with $L \equiv 1$. Hence, showing that R_T converges in probability for all bounded loss functions yields the result. Now,

$$\begin{aligned}
\bar{L}^{-1} R_T &\leq \int\int \left| \Theta_u \mathcal{V}_T(u) \frac{w(\theta_0 + T^{-1/2}u) LR_T(u, \delta)}{D} - w(\theta_0) \frac{\widehat{LR}_T(u, \delta)}{\widehat{D}} \right| dQ(\delta) du \\
&\leq D^{-1} \int\int \left| \Theta_u \mathcal{V}_T(u) w(\theta_0 + T^{-1/2}u) LR_T(u, \delta) - w(\theta_0)\, \widehat{LR}_T(u, \delta) \right| dQ(\delta) du \\
&\quad + w(\theta_0) \left| D^{-1} - \hat{D}^{-1} \right| \int\int \widehat{LR}_T(u, \delta)\, dQ(\delta) du.
\end{aligned}$$

We show that

$$N_T = \int\int \left| \Theta_u \mathcal{V}_T(u) w(\theta_0 + T^{-1/2}u) LR_T(u, \delta) - w(\theta_0)\, \widehat{LR}_T(u, \delta) \right| dQ(\delta) du = o_p(1),$$

$\widehat{D}^{-1} = O_p(1)$, and $\widehat{D} = O_p(1)$, which yields as a corollary $D - \hat{D} = o_p(1)$ and $R_T = o_p(1)$.

Write

$$\int\int \widehat{LR}_T(u, \delta)\, dQ(\delta) du = \int \widehat{LR}_T(0, \delta)\, dQ(\delta) \int \exp\left(\hat{u}'\hat{H}u - \tfrac{1}{2}u'\hat{H}u\right) du$$

and note that $\int \exp\left(\hat{u}'\hat{H}u - \frac{1}{2}u'\hat{H}u\right)du = \exp\left(\frac{1}{2}\hat{u}'\hat{H}\hat{u}\right)\left|\hat{H}\right|^{-1/2}(2\pi)^{k/2}$. Given the consistency of $\hat{\theta}$ under θ_0 stable (see Lemma (2)), $\frac{1}{2}\hat{u}'\hat{H}^{-1}\hat{u} = O_p(1)$, and $\hat{H} - H = o_p(1)$. Since H is not singular by (LLLN), we can conclude that both $\exp\left(\frac{1}{2}\hat{u}'\hat{H}^{-1}\hat{u}\right)\left|\hat{H}\right|^{-1/2}$ and $\exp\left(-\frac{1}{2}\hat{u}'\hat{H}^{-1}\hat{u}\right)\left|\hat{H}\right|^{1/2}$ are $O_p(1)$. Lemma (6) states that both $\int \widehat{LR}_T(0,\delta)\,dQ(\delta)$ and $\left(\int \widehat{LR}_T(0,\delta)\,dQ(\delta)\right)^{-1}$ are $O_p(1)$. Both $\widehat{D}^{-1}$ and $\widehat{D}$ are therefore $O_p(1)$.

It is now shown that $N_T = o_p(1)$. Let $\mathcal{S}_T = \{\delta : T^{1/2}\sup_{i,t}|\delta_{i,t}| < T^{1/4}\}$,

$$\begin{aligned}
N_{B,T} &= \int\int \mathcal{S}_T\left|\mathcal{V}_T(u)\Theta_u LR_T(u,\delta)w(\theta_0 + T^{-1/2}u) - w(\theta_0)\widehat{LR}_T(u,\delta)\right| dQ(\delta)du,\\
N_{G,T} &= \int_{||u||<a_T}\int \mathcal{S}_T\left|\mathcal{V}_T(u)\Theta_u LR_T(u,\delta)w(\theta_0 + T^{-1/2}u) - w(\theta_0)\widehat{LR}_T(u,\delta)\right| dQ(\delta)du,
\end{aligned}$$

where $(a_T)_{T\geq 1}$ is some increasing sequence to be defined below. Then

$$N_T = (N_T - N_{B,T}) + (N_{B,T} - N_{G,T}) + N_{G,T},$$

the sum of three terms, each of which is shown to converge in probability. From now on, $E_\delta[\cdot]$ is shorthand notation for integration on $\mathbb{R}^{Tk}$ with respect to Q.

For the first term, note that, by Markov's inequality and because the tail probability of the supremum of a Gaussian random walk decays exponentially,

$$\begin{aligned}
&P\Big(\int_{\Theta_u}\int_{\mathcal{V}_T(u)}(1-\mathcal{S}_T)w(\theta_0 + T^{-1/2}u)LR_T(u,\delta)dQ(\delta)du > \eta\Big)\\
\leq\;& \eta^{-1}\int\int_{\Theta_u}\int_{\mathcal{V}_T(u)}(1-\mathcal{S}_T)w(\theta_0 + T^{-1/2}u)LR_T(u,\delta)dQ(\delta)du f_T(\theta_0,0)d\mu_T\\
=\;& \eta^{-1}\int_{\Theta_u}\int_{\mathcal{V}_T(u)}\left(\int f_T(\theta_0 + T^{-1/2}u,\delta)d\mu_T\right)(1-\mathcal{S}_T)w(\theta_0 + T^{-1/2}u)dQ(\delta)du\\
\leq\;& \eta^{-1}E_\delta(1-\mathcal{S}_T)\int_{\Theta_u} w(\theta_0 + T^{-1/2}u)du\\
=\;& \eta^{-1}E_\delta(1-\mathcal{S}_T)O\left(T^{k/2}\right)\\
\to\;& 0.
\end{aligned}$$

By Lemma (6),

$$\begin{aligned}
&\int\int(1-\mathcal{S}_T)w(\theta_0)\widehat{LR}_T(u,\delta)dQ(\delta)du\\
\leq\;& w(\theta_0)\left[E_\delta(1-\mathcal{S}_T)\right]^{1/2}\int\left[\int \widehat{LR}_T(u,\delta)^2 dQ(\delta)\right]^{1/2}du\\
=\;& w(\theta_0)\left[E_\delta(1-\mathcal{S}_T)\right]^{1/2}O_p(1) \xrightarrow{p} 0.
\end{aligned}$$

Hence, $(N_T - N_{B,T}) = o_p(1)$.

For the second term, pick any sequence of decreasing numbers $\tilde{a}_i$ such that $\tilde{a}_i \to 0$ as $i \to \infty$. For each i, by assumption (ID), there exists T_i^* such that for all $T > T_i^*$

$$P\Big(\sup_{\theta\in\Theta,||\theta-\theta_0||\geq\tilde{a}_i} T^{-1}\sum\sup_{||v||<T^{1/4},\theta+T^{-1/2}v\in\Theta}(l_t(\theta + T^{-1/2}v) - l_t(\theta_0)) < -K(\tilde{a}_i)\Big) \geq 1 - \tilde{a}_i$$

For any T, let i_T be the largest i such that (i) $T > T_i^*$, (ii) $T^{1/2}K(\tilde{a}_i) > 1$ and (iii) $T^{1/4}\tilde{a}_i > 1$. Note that $i_T \to \infty$. Set $a_T = T^{1/2}\tilde{a}_{i_T}$. Then

$$P(\sup_{\theta\in\Theta, T^{1/2}||\theta-\theta_0||\geq a_T} T^{-1}\sum \sup_{||v||<T^{1/4}} (l_t(\theta+T^{-1/2}v) - l_t(\theta_0)) < -K(T^{-1/2}a_T)) \geq 1 - T^{-1/2}a_T \to 1$$

and $T^{1/2}K(T^{-1/2}a_T) > 1$, and $a_T \to \infty$.

Then

$$\begin{aligned}
&\int_{\Theta_u\cap\{u:||u||\geq a_T\}}\int_{\mathcal{V}_T(u)} \mathcal{S}_T w(\theta_0+T^{-1/2}u)LR_T(u,\delta)dQ(\delta)du \\
\leq\ & \exp\left[\sup_{\theta\in\Theta,T^{1/2}||\theta-\theta_0||\geq a_T}\ \sup_{\delta\in\mathcal{S}_T\cap\mathcal{V}_T(\theta)}\sum(l_t(\theta+\delta_t)-l_t(\theta_0))\right]\int w(\theta_0+T^{-1/2}u)du \\
=\ & \exp\left[\sup_{\theta\in\Theta,T^{1/2}||\theta-\theta_0||\geq a_T}\sum\ \sup_{||v||<T^{-1/4},\theta+T^{-1/2}v\in\Theta}(l_t(\theta+v)-l_t(\theta_0))\right]O\left(T^{k/2}\right) \\
\leq\ & \exp[-TK(T^{-1/2}a_T)]O\left(T^{k/2}\right) \leq \exp[-T^{1/2}]O\left(T^{k/2}\right) \to 0
\end{aligned}$$

with probability converging to one. Moreover,

$$\begin{aligned}
&\int_{||u||\geq a_T}\int \mathcal{S}_T\widehat{LR}_T w(\theta_0)dQ(\delta)du \\
\leq\ & w(\theta_0)\int_{||u||\geq a_T}\exp[\hat{u}'\hat{H}u - \tfrac{1}{2}u'\hat{H}u]du\int\widehat{LR}_T(0,\delta)dQ(\delta).
\end{aligned}$$

By Lemma (6) again, $\int\widehat{LR}_T(0,\delta)dQ(\delta) = O_p(1)$; with $Z \sim N(0, I_k)$,

$$\begin{aligned}
\int_{||u||\geq a_T}\exp\left(\hat{u}'\hat{H}u - \tfrac{1}{2}u'\hat{H}u\right)du &= \exp\left(\frac{1}{2}\hat{u}'\hat{H}\hat{u}\right)\left|\hat{H}\right|^{-1/2}(2\pi)^{k/2} \\
&\quad\times\left|\hat{H}\right|^{1/2}(2\pi)^{-k/2}\int_{||u||\geq a_T}\exp\left(-\tfrac{1}{2}(u-\hat{u})'\hat{H}(u-\hat{u})\right)du \\
&= O_p(1)\,P\left(\left\|\hat{u}+\hat{H}^{-1/2}Z\right\|\geq a_T\right) \\
&\leq O_p(1)\,P\left(\|Z\|\geq(a_T-\|\hat{u}\|)\left\|\hat{H}^{1/2}\right\|\right)\xrightarrow{p}0
\end{aligned}$$

since $(a_T - \|\hat{u}\|)\left\|\hat{H}^{1/2}\right\| = a_T\left\|H^{1/2}\right\| + O_p(1) \xrightarrow{p} \infty$ as $T \to \infty$. Hence,

$$\begin{aligned}
N_{B,T} - N_{G,T} \leq\ & \int_{\Theta_u\cap\{u:||u||\geq a_T\}}\int_{\mathcal{V}_T(u)}\mathcal{S}_T LR_T(u,\delta)w(\theta_0+T^{-1/2}u)dQ(\delta)du \\
& +\int_{||u||\geq a_T}\int\mathcal{S}_T\widehat{LR}_T w(\theta_0)dQ(\delta)du \xrightarrow{p} 0.
\end{aligned}$$

Let $\mathcal{A}_T := \left\{v\in\mathbb{R}^k : \|v\| < a_T\right\}$ and note that for T large enough, $\mathcal{A}_T \subset \Theta_u$, by assumption (DIFF), as θ_0 is an interior point of Θ. By the same argument, for large enough T, $\mathcal{S}_T \subset \mathcal{V}_T(u)$ for all $u \in \mathcal{A}_T$. Assume in the following that T is large enough for these two set inclusions to hold.

Since $a_T = o\left(T^{1/2}\right)$, $T^{-1/2}u \to 0$ for all $u \in \mathcal{A}_T$, so that by the continuity of w at θ_0,

$$\sup_{||u||<a_T} \left| w(\theta_0 + T^{-1/2}u) - w\left(\theta_0\right) \right| = o(1).$$

Moreover, it will be shown that $\int_{\mathcal{A}_T} \int \mathcal{S}_T \left| LR_T(u,\delta) - \widehat{LR}_T(u,\delta) \right| dQ(\delta)du = o_p(1)$; as argued above, $\int \int \widehat{LR}_T\left(u,\delta\right) dQ(\delta)du = O_p\left(1\right)$. It follows that $\int_{\mathcal{A}_T} \int \mathcal{S}_T LR_T(u,\delta)dQ(\delta)du \leq \int \int \widehat{LR}_T(u,\delta)dQ(\delta)du + o_p\left(1\right)$, so that

$$\begin{aligned} & \sup_{u \in \mathcal{A}_T} \left| w(\theta_0 + T^{-1/2}u) - w\left(\theta_0\right) \right| \int_{\mathcal{A}_T} \int \mathcal{S}_T LR_T(u,\delta)dQ(\delta)du \\ \leq \; & o(1)O_p(1) + o_p(1), \end{aligned}$$

which gives the following bound:

$$\begin{aligned} N_{G,T} &= \int \int \mathcal{A}_T \mathcal{S}_T \left| \mathcal{V}_T(u)\Theta_u LR_T(u,\delta) w(\theta_0 + T^{-1/2}u) - \widehat{LR}_T w(\theta_0) \right| dQ(\delta)du \\ &\leq w(\theta_0) \int \int \mathcal{A}_T \mathcal{S}_T \left| LR_T(u,\delta) - \widehat{LR}_T(u,\delta) \right| dQ(\delta)du \\ &\quad + \sup_{u \in \mathcal{A}_T} \left| w(\theta_0 + T^{-1/2}u) - w\left(\theta_0\right) \right| \int \int \mathcal{A}_T \mathcal{S}_T LR_T(u,\delta)dQ(\delta)du \\ &= o_p(1) \end{aligned}$$

The remainder of the proof now shows that $\int \int \mathcal{A}_T \mathcal{S}_T \left| LR_T(u,\delta) - \widehat{LR}_T(u,\delta) \right| dQ(\delta)du = o_p(1)$.

On $\mathcal{A}_T$ and $\mathcal{S}_T$, $T^{-1/2}u = o\left(1\right)$ and $\sup_t |\delta_t| = o\left(1\right)$, so that as $T \to \infty$, $\theta_0 + T^{-1/2}u + \delta_t \in \Theta_0 \forall t$ and hence $l_t(\theta)$ is almost surely twice differentiable with respect to θ by assumption (DIFF). Also $\hat{\theta} \in \Theta_0$ with probability converging to one. A series of second order Taylor expansions gives the following expression for $LR_T(u,\delta)$ (on $\mathcal{A}_T$ and $\mathcal{S}_T$ and for T large enough):

$$\begin{aligned} LR_T(u,\delta) &= \frac{f_T(\theta_0 + T^{-1/2}u,\delta)}{f_T(\theta_0,0)} = \frac{f_T(\theta_0 + T^{-1/2}u,\delta)}{f_T(\theta_0 + T^{-1/2}u,0)} \frac{f_T(\theta_0 + T^{-1/2}u,0)}{f_T(\theta_0,0)} \\ &= \exp[\sum s_t(\theta_0 + T^{-1/2}u)'\delta_t + \tfrac{1}{2} \sum \delta_t' h_t(\tilde{\theta}_t)\delta_t] \\ &\qquad \exp[T^{-1/2} \sum s_t(\theta_0)'u + \tfrac{1}{2} u' \left(T^{-1} \sum h_t(\tilde{\theta}^1) \right) u] \\ &= \exp[\sum s_t(\theta_0)'\delta_t + \tfrac{1}{2} \sum \delta_t' h_t(\tilde{\theta}_t)\delta_t + T^{-1/2}u' \sum h_t(\bar{\theta})\delta_t] \\ &\qquad \exp[T^{-1/2} \sum s_t(\theta_0)'u + \tfrac{1}{2} u' \left(T^{-1} \sum h_t(\tilde{\theta}^1) \right) u] \\ &= \exp[\sum s_t(\hat{\theta})'\delta_t + T^{-1/2}\hat{u}' \sum h_t(\tilde{\theta}^2)\delta_t + \tfrac{1}{2} \sum \delta_t' h_t(\tilde{\theta}_t)\delta_t + T^{-1/2}u' \sum h_t(\bar{\theta})\delta_t] \\ &\qquad \exp[\hat{u}'\tilde{H}u + \tfrac{1}{2} u' \left(T^{-1} \sum h_t(\tilde{\theta}^1) \right) u], \end{aligned}$$

where $\tilde{\theta}_t$ lies on the line segment between $\theta_0 + T^{-1/2}u$ and $\theta_0 + T^{-1/2}u + \delta_t$, $\tilde{\theta}^1$ and $\bar{\theta}$ on that between θ_0 and $\theta_0 + T^{-1/2}u$, $\tilde{\theta}^2$ on that between $\hat{\theta}$ and θ_0, and $\tilde{H} = -T^{-1} \sum h_t(\tilde{\theta}^2)$ if it is positive definite, otherwise $\tilde{H} = I_k$. Define

$$
\begin{aligned}
\varsigma_T &= T^{-1/2}\hat{u}' \sum h_t(\tilde{\theta}^2)\delta_t + \tfrac{1}{2}\sum \delta_t' \left(h_t(\tilde{\theta}_t) + \hat{H} \right) \delta_t + T^{-1/2} u' \sum h_t(\bar{\theta})\delta_t \\
&\quad + \hat{u}'(\tilde{H} - \hat{H})u + \tfrac{1}{2}u' \left(T^{-1} \sum h_t(\tilde{\theta}^1) + \hat{H} \right) u \\
\overline{LR}_T(u,\delta) &= LR_T(u,\delta)\exp[-\hat{u}'\hat{H}u + \tfrac{1}{2}u'\hat{H}u] \\
&= \exp\left(\sum s_t(\hat{\theta})'\delta_t - \frac{1}{2}\sum \delta_t' \hat{H}\delta_t + \varsigma_T \right) = \widehat{LR}_T(0,\delta)\exp \varsigma_T
\end{aligned}
$$

and let $\Phi(u)$ stand for the distribution of $u \sim N(\hat{u}, \hat{H}^{-1})$. Then

$$
\begin{aligned}
&\int_{||u||<a_T} \int \mathcal{S}_T |\widehat{LR}_T(u,\delta) - LR_T(u,\delta)| dQ(\delta) du \\
= \; & (2\pi)^{k/2} |\hat{H}|^{-1/2} \exp[\tfrac{1}{2}\hat{u}'\hat{H}\hat{u}] \int_{||u||<a_T} \int \mathcal{S}_T |\widehat{LR}_T(0,\delta) - \overline{LR}_T(u,\delta)| dQ(\delta) d\Phi(u).
\end{aligned}
$$

As argued above, $(2\pi)^{k/2}|\hat{H}|^{-1/2}\exp[\frac{1}{2}\hat{u}'\hat{H}\hat{u}] = O_p(1)$. Also

$$
\begin{aligned}
&\left[\int_{||u||<a_T} \int \mathcal{S}_T |\widehat{LR}_T(0,\delta) - \overline{LR}_T(u,\delta)| dQ(\delta) d\Phi(u) \right]^2 \\
\leq \; & \int_{||u||<a_T} \int \mathcal{S}_T |\widehat{LR}_T(0,\delta) - \overline{LR}_T(u,\delta)|^2 dQ(\delta) d\Phi(u) \\
= \; & \int_{||u||<a_T} \int \mathcal{S}_T [2\widehat{LR}_T(0,\delta)(\widehat{LR}_T(0,\delta) - \overline{LR}_T(u,\delta)) \\
& + \overline{LR}_T(u,\delta)^2 - \widehat{LR}_T(0,\delta)^2] dQ(\delta) d\Phi(u).
\end{aligned}
$$

Now,

$$
\begin{aligned}
E_\delta \mathcal{S}_T \widehat{LR}_T(0,\delta)(\widehat{LR}_T(0,\delta) - \overline{LR}_T(u,\delta)) &= E_\delta \mathcal{S}_T \widehat{LR}_T(0,\delta)^2 \left(1 - \exp \varsigma_T\right) \\
&\leq E_\delta \mathcal{S}_T \exp\left(2\sum s_t(\hat{\theta})'\delta_t \right) (1 - \exp\varsigma_T) \\
&\leq \left(E_\delta \mathcal{S}_T \exp 4 \sum s_t(\hat{\theta})'\delta_t \right)^{1/2} \left(E_\delta \mathcal{S}_T (1 - \exp\varsigma_T)^2 \right)^{1/2}.
\end{aligned}
$$

(where the first inequality follows from the negative definiteness of $-\hat{H}$) and

$$
\begin{aligned}
E_\delta \mathcal{S}_T \left(\widehat{LR}_T(0,\delta)^2 - \overline{LR}_T(u,\delta)^2 \right) &\leq E_\delta \mathcal{S}_T \exp\left(2\sum s_t(\hat{\theta})'\delta_t \right) (1 - \exp 2\varsigma_T) \\
&\leq \left(E_\delta \mathcal{S}_T \exp 4 \sum s_t(\hat{\theta})'\delta_t \right)^{1/2} \left(E_\delta \mathcal{S}_T (1 - \exp 2\varsigma_T)^2 \right)^{1/2}.
\end{aligned}
$$

Essentially same problem: show that $E_\delta \mathcal{S}_T \exp 4 \sum s_t(\hat{\theta})'\delta_t = O_p(1)$ and

$$
\left(\int \mathcal{A}_T \left(E_\delta \mathcal{S}_T (1 - \exp \tilde{p}\varsigma_T)^2 \right)^{1/2} d\Phi(u) \right)^2 \leq E_u \mathcal{A}_T E_\delta \mathcal{S}_T (1 - \exp\tilde{p}\varsigma_T)^2 = o_p(1)
$$

for $\tilde{p} = 1, 2$. For simplicity, set $\tilde{p} = 1$; for $\tilde{p} = 2$, the proof is essantially the same.

First, recall that $\hat{u} = T^{1/2}\left(\hat{\theta} - \theta_0\right) = \tilde{H}^{-1}T^{-1/2}\sum s_t(\theta_0)$. Hence,

$$\begin{aligned} T^{-1/2}\sum\nolimits_{t=1}^{[\lambda T]}\left(s_t(\hat{\theta}) + \tilde{H}T^{-1/2}\hat{u}\right) &= T^{-1/2}\sum\nolimits_{t=1}^{[\lambda T]}\left(s_t(\theta_0) + \left(h_t(\tilde{\theta}^2) + \tilde{H}\right)T^{-1/2}\hat{u}\right) \\ &= T^{-1/2}\sum\nolimits_{t=1}^{[\lambda T]} s_t(\theta_0) \\ &\quad + T^{-1}\sum\nolimits_{t=1}^{[\lambda T]}\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)\left(\tilde{H}^{-1}T^{-1/2}\sum s_t(\theta_0)\right) \end{aligned}$$

where the process $\lambda \mapsto T^{-1}\sum_{t=1}^{[\lambda T]}\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)\left(\tilde{H}^{-1}T^{-1/2}\sum s_t(\theta_0)\right)$ converges to zero uniformly in λ, given conditions (LLLN) and (FCLT). Hence, $\lambda \mapsto T^{-1/2}\sum_{t=1}^{[\lambda T]} s_t(\hat{\theta}) + \frac{[\lambda T]}{T}\tilde{H}\hat{u}$ and $\lambda \mapsto T^{-1/2}\sum_{t=1}^{[\lambda T]} s_t(\theta_0)$ converge to the same limiting process. Because of the normalization constraint on δ, $\sum_t\left(T^{-1}\sum_{l=1}^{T} s_l'(\theta_0)\right)\delta_t = 0$, and we have

$$E_\delta \mathcal{S}_T \exp\left(4\sum\nolimits_t s_t(\hat{\theta})'\delta_t\right) \leq E_\delta \exp\left(4\sum\nolimits_t\left(s_t(\hat{\theta}) + T^{-1/2}\tilde{H}\hat{u}\right)'\delta_t\right),$$

to which one can apply Lemma (1), and one obtains $E_\delta \mathcal{S}_T \exp 4\sum s_t(\hat{\theta})'\delta_t = O_p(1)$.

Next, we bound $E_\delta\mathcal{S}_T(1-\exp(\varsigma_T))^2 \leq 1 + E_\delta\mathcal{S}_T\exp(2\varsigma_T) - 2E_\delta\mathcal{S}_T\exp(\varsigma_T)$. Let $b_T := T^{-1/2}(1+a_T)$ and $\mathcal{B}_T := \{\theta : \|\theta - \theta_0\| \leq b_T\}$, a decreasing sequence of neighborhoods of θ_0. On $\mathcal{S}_T$ and $\mathcal{A}_T$, $\tilde{\theta}_t$, $\tilde{\theta}^1$ and $\bar{\theta}$ are in $\mathcal{B}_T$, and, by the $\sqrt{T}$-consistency of $\hat{\theta}$, $\hat{\theta}$ and $\tilde{\theta}^2$ are in $\mathcal{B}_T$ with probability approaching one. For each t define

$$c_{T,t}(\theta_0) = \sum\nolimits_{i,j\leq k}\sup_{\theta\in\mathcal{B}_T}\left|\left(h_t(\theta) - h_t(\theta_0)\right)_{i,j}\right|,$$

a random variable that does not depend on δ or u. For any $k\times k$ symmetric matrix B, the ith eigenvalue of which we denote λ_i, and any $v\in\mathbb{R}^k$, $|v'Bv| \leq (\sup_{i\leq k}|\lambda_i|)v'v$, and $\sup_{i\leq k}|\lambda_i| \leq \sqrt{tr\,(B'B)} \leq \sum_{i,j}|b_{i,j}|$. Hence, on $\mathcal{S}_T$ and $\mathcal{A}_T$,

$$\sum \delta_t'\left(h_t(\tilde{\theta}_t) + \hat{H}\right)\delta_t \leq \sum \delta_t'\left(h_t(\theta_0) + \hat{H} + c_{T,t}(\theta_0)\,k^2 I_k\right)\delta_t.$$

Using the normalization constraint on δ again, $\sum_t\left(T^{-1/2}\hat{u}'\tilde{H}\right)\delta_t = 0$ and $\sum_t\left(T^{-1/2}u'\tilde{H}\right)\delta_t = 0$, and so

$$\begin{aligned} & E_\delta\mathcal{S}_T\exp(2\varsigma_T) \\ \leq\ & E_\delta\exp 2[T^{-1/2}\hat{u}'\sum\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)\delta_t + \frac{1}{2}\sum\delta_t'(h_t(\theta_0) + \hat{H} + c_{T,t}(\theta_0)\,I_k)\delta_t \\ & + T^{-1/2}u'\sum\left(h_t(\bar{\theta}) + \tilde{H}\right)\delta_t + \hat{u}'(\tilde{H} - \hat{H})u + \frac{1}{2}u'\left(T^{-1}\sum h_t(\theta_0) + \hat{H} + c_{T,t}(\theta_0)\,I_k\right)u] \end{aligned}$$

The same procedure with now $c_{T,t}(\theta_0) := \sum_{i,j\leq k}\inf_{\theta\in\mathcal{B}_T}\left|\left(h_t(\theta) - h_t(\theta_0)\right)_{i,j}\right|$ gives an upper bound for $-2E_\delta\mathcal{S}_T\exp(\varsigma_T)$. In both cases, the bound takes the form

$$\begin{aligned} & \exp[p\left(\hat{u}'(\tilde{H} - \hat{H})u + \frac{1}{2}u'\left(T^{-1}\sum h_t(\theta_0) + \hat{H} + c_{T,t}(\theta_0)\,I_k\right)u\right)] \\ & \times E_\delta\exp\left(\sum v_t'\delta_t + \frac{1}{2}\sum\delta_t' w_t\delta_t + \sum u'\tilde{w}_t\delta_t\right), \end{aligned}$$

with $p=1,2$ and where $v_t' = pT^{-1/2}\hat{u}'\left(h_t(\tilde{\theta}^2)+\tilde{H}\right)$, $w_t = p\left(h_t(\theta_0)+\hat{H}+c_{T,t}\left(\theta_0\right)\right)$, and $\tilde{w}_t = pT^{-1/2}\left(h_t(\bar{\theta})+\tilde{H}\right)$, which satisfy the assumption of Lemma (3):

$$\begin{aligned}\sup_\lambda T^{-1}\sum_{t=1}^{[\lambda T]}\left(h_t(\theta_0)+\hat{H}+c_{T,t}\left(\theta_0\right)I_k\right) &= o_p\left(1\right),\\ \sup_\lambda T^{-1}\sum_{t=1}^{[\lambda T]}\left(h_t(\bar{\theta})+\tilde{H}\right) &= o_p\left(1\right),\\ \sup_\lambda T^{-1}\sum_{t=1}^{[\lambda T]}\left(h_t(\tilde{\theta}^2)+\tilde{H}\right)\hat{u} &= o_p\left(1\right).\end{aligned}$$

An application of Lemma (3) thus gives

$$E_\delta \exp\left(\sum v_t'\delta_t+\frac{1}{2}\sum \delta_t' w_t\delta_t+\sum u'\tilde{w}_t\delta_t\right) = \left(1+\tilde{D}\right)\exp u'\tilde{\Delta}_{p,T}u$$

where, for $p=1,2$, $\hat{\Delta}_{p,T}$ are $o_p(1)$ $k\times k$ matrices that do not depend on u, and $\sup_u \tilde{D} = o_p\left(1\right)$. Hence

$$\begin{aligned}&\left(-2\right)^{2-p}E_\delta\mathcal{S}_T\exp\left(p\varsigma_T\right)\\ \leq\ &\left(-2\right)^{2-p}\exp[u'\left(\tilde{\Delta}_{p,T}+pT^{-1}\left(\sum h_t(\theta_0)+\hat{H}+c_{T,t}\left(\theta_0\right)I_k\right)\right)u+p\hat{u}'(\tilde{H}-\hat{H})u](1+\tilde{D})\\ =\ &\left(-2\right)^{2-p}\exp[u'\Delta_{p,T}u+p\hat{u}'(\tilde{H}-\hat{H})u](1+\tilde{D})\end{aligned}$$

where $\Delta_{p,T}:=\hat{\Delta}_{p,T}+pT^{-1}\left(\sum h_t(\theta_0)+\hat{H}+c_{T,t}\left(\theta_0\right)I_k\right)=o_p\left(1\right)$.

Now, with probability approaching one, $\hat{H}-2\Delta_{p,T}$ is non-singular, and so

$$\begin{aligned}&E_u\exp[u'\Delta_{p,T}u+p\hat{u}'(\tilde{H}-\hat{H})u](1+\tilde{D})\\ =\ &E_u\exp[u'\Delta_{p,T}u+p\hat{u}'(\tilde{H}-\hat{H})u]+o_p\left(1\right)\\ =\ &o_p\left(1\right)+|\hat{H}^{1/2}||\hat{H}-2\Delta_{p,T}|^{-1/2}\\ &\times\exp\left(-\frac{1}{2}\hat{u}'\hat{H}\hat{u}+\frac{1}{2}\hat{u}'(p(\tilde{H}-\hat{H})+\hat{H})\left(\hat{H}-2\Delta_{p,T}\right)^{-1}(p(\tilde{H}-\hat{H})+\hat{H})\hat{u}\right)\\ =\ &o_p\left(1\right)+\left(1+o_p\left(1\right)\right)\exp\left(-\frac{1}{2}\tilde{Z}'H\tilde{Z}+\frac{1}{2}\tilde{Z}'HH^{-1}H\tilde{Z}+o_p\left(1\right)\right)\\ =\ &1+o_p\left(1\right)\end{aligned}$$

(where $\tilde{Z}$ is a random variable such that $\hat{u}\xrightarrow{d}\tilde{Z}$) and $E_u\left(1-\mathcal{A}_T\right)\exp[u'\Delta_{p,T}u+p\hat{u}'(\tilde{H}-\hat{H})u](1+\tilde{D})\leq\left(E_u\left(1-\mathcal{A}_T\right)\right)^{1/2}O_p\left(1\right)\xrightarrow{p}0$. Hence

$$\begin{aligned}&E_u\mathcal{A}_TE_\delta\mathcal{S}_T(1-\exp\varsigma_T)^2\\ \leq\ &E_u\mathcal{A}_T((1+E_\delta\mathcal{S}_T\exp 2\varsigma_T-2E_\delta\mathcal{S}_T\exp\varsigma_T)\\ \leq\ &1-2E_u\exp[u'\Delta_{1,T}u+\hat{u}'(\tilde{H}-\hat{H})u]+E_u\exp[u'\Delta_{2,T}u+2\hat{u}'(\tilde{H}-\hat{H})u]+o_p\left(1\right)\\ =\ &1-2|\hat{H}^{1/2}||\hat{H}-2\Delta_{2,T}|^{-1/2}\left(1+o_p\left(1\right)\right)+|\hat{H}^{1/2}||\hat{H}-2\Delta_{1,T}|^{-1/2}\left(1+o_p\left(1\right)\right)+o_p\left(1\right)\\ =\ &o_p\left(1\right)\end{aligned}$$

Summing up, $\int_{||u||<a_T}\int \mathcal{S}_T|\widehat{LR}_T(u,\delta)-LR_T(u,\delta)|dQ(\delta)du = o_p(1)$, which proves that $N_{G,T}=o_p(1)$, which proves part (c) of the Theorem.

Proof of part (a). Define $f_1=\int_\Theta\int_{\mathcal{V}_T(\theta)} w(\theta)f_T(\theta,\delta)dQ(\delta)d\theta$ so that

$$WAR(\hat{a})=\int\frac{\int_\Theta\int_{\mathcal{V}_T(\theta)} w(\theta)LR(\theta,\delta)\,L(\theta,\delta,\hat{a})dQ(\delta)d\theta}{\int_\Theta\int_{\mathcal{V}_T(\theta)} w(\theta)LR(\theta,\delta)\,dQ(\delta)d\theta}f_1 d\mu_T.$$

For the decision rule $\hat{a}$, write

$$\begin{aligned}\widehat{WAR}(\hat{a}) &= \int\int\int L(\hat{\theta}+T^{-1/2}\hat{H}^{-1/2}u,\Sigma\hat{s}+(\Sigma^{1/2})^{+}\delta,\hat{a})d\Phi_{Tk}(\delta)d\Phi_k(u)f_1 d\mu_T\\ &= \int\left(\frac{\int\int w(\theta_0)\,\widehat{LR}_T(\theta,\delta)L(\theta,\delta,a)}{w(\theta_0)\int\int\widehat{LR}_T(\theta,\delta)dQ(\delta)d\theta}dQ(\delta)d\theta\right)f_1 d\mu_T\end{aligned}$$

and note that

$$\begin{aligned}WAR(\hat{a})-WAR(\hat{a}^*) &= \left(\widehat{WAR}(\hat{a})-\widehat{WAR}(\hat{a}^*)\right)\\ &\quad+\left(WAR(\hat{a})-\widehat{WAR}(\hat{a})\right)+\left(\widehat{WAR}(\hat{a}^*)-WAR(\hat{a}^*)\right)\\ &\geq \left(WAR(\hat{a})-\widehat{WAR}(\hat{a})\right)+\left(\widehat{WAR}(\hat{a}^*)-WAR(\hat{a}^*)\right)\end{aligned}$$

and the result is proved if $\underline{\lim}_{T\to\infty}[WAR(\hat{a})-\widehat{WAR}(\hat{a})]=0$ for all decision rules $\hat{a}$. With R_T defined as in the proof of part (c),

$$\begin{aligned}\left|WAR(\hat{a})-\widehat{WAR}(\hat{a})\right| &\leq \int R_T f_1 d\mu_T\\ &\leq \int_\Theta w(\theta)\left(\int R_T\left(\int f_T(\theta,\delta)dQ(\delta)\right)d\mu_T\right)d\theta\end{aligned}$$

The latter expression converges to zero if $\int R_T\left(\int f_T(\theta,\delta)dQ(\delta)\right)d\mu_T$ is bounded uniformly in θ and converges to zero for all $\theta\in\Theta$ such that $w(\theta)>0$. But $f_1^*(\theta)=\int f_T(\theta,\delta)dQ(\delta)$ is a density for the data y_T; Lemma (9) shows that it is contiguous to $f_T(\theta,0)$, for all such θ. Hence, given that $R_T\leq 2\bar{L}$, if R_T converges to zero in probability under θ stable, $\forall\eta>0$, $\exists T^*: P(R_T>\eta)\leq\frac{1}{2}\bar{L}^{-1}\frac{\eta}{2}$ and

$$\int R_T f_T(\theta_0,0)d\mu_T\leq P(R_T>\eta/2)\,2\bar{L}+P(R_T\leq\eta/2)\,\frac{\eta}{2}\leq\eta$$

for all $T\geq T^*$ and, together with contiguity, $\int R_T f_1^*(\theta)d\mu_T\to 0$ for all θ, which completes the proof of part (a).

Proof of part (b). Write $WAP(\varphi_T)=\int\varphi_T\left(\int_{\mathcal{V}_T(\theta)} f_T(\theta,\delta)dQ(\delta)\right)d\mu_T=\kappa\int\varphi_T\tilde{f}(\theta)d\mu_T$, where $\tilde{f}(\theta)=\kappa^{-1}\int_{\mathcal{V}_T(\theta)} f_T(\theta,\delta)dQ(\delta)$ and κ is a normalisation factor so that $\tilde{f}(\theta)$ is a density. By the Neyman-Pearson Lemma, the efficient test is based on the likelihood ratio $LR_T(y_T,\theta)=\tilde{f}(\theta)/f(\theta,0)$ which is shown in Lemma (9) to be close to $\int\widehat{LR}_T(\theta,\delta)\,dQ(\delta)$. By Lemma (8), with $\upsilon=\hat{s}$ and $C=\hat{H}$, $\int\widehat{LR}_T(\theta,\delta)\,dQ(\delta)=\prod_{i=1}^{k}\left(\frac{1-r_{a_i}^{2T}}{T(1-r_{a_i}^2)r_{a_i}^{T-1}}\right)^{-1/2}\exp\left(\frac{1}{2}\hat{s}\Sigma\hat{s}\right)$, the asymptotic distribution of which is independent of θ_0 under the null $H_0:\delta=0$

Proof of Theorem 2:
to be added.

Lemma 1 *For any $k \times 1$ process (z_t) for which a FCLT holds,*

$$E_\delta \exp\left(\sum\nolimits_{t=1}^{T} z_t' \delta_t\right) = O_p(1).$$

Proof. Let $z_T := (z_1', ..., z_T')'$ and recall that $\delta \overset{d}{=} \left(T^{-1} M_e F \otimes \Omega^{1/2}\right) \varepsilon$, where $\varepsilon \sim N(0, I_{kT})$, so that, with $\widetilde{z}_T := \left(T^{-1/2} F' M_e \otimes \Omega^{1/2}\right) z_T$,

$$\begin{aligned} E_\delta \exp\left(\sum\nolimits_{t=1}^{T} z_t' \delta_t\right) &= E_\varepsilon \exp\left(T^{-1/2} \widetilde{z}_T' \varepsilon\right) \\ = \prod\nolimits_{i \leq k} \exp\left(T^{-1} \sum_t \widetilde{z}_{i,t}^2\right) &= O_p(1). \end{aligned}$$

■

Lemma 2 *Under Condition (2) and θ_0 stable, the maximum likelihood estimator of $\hat{\theta}$ is consistent for θ_0 and $\hat{H}$ provides a consistent estimator for H.*

Proof. to be added. ■

Lemma 3 *Let $\{w_{T,t}\}_{t=1}^T$, $\{\tilde{w}_{T,t}\}_{t=1}^T$ be sequences of $k \times k$ matrices and $\{v_{T,t}\}_{t=1}^T$ be a sequence of $k \times 1$ vectors satisfying*

$$\sup_\lambda T^{-1/2} \sum_{t=1}^{[\lambda T]} (T^{-1/2} w_{T,t}, \tilde{w}_{T,t}, v_{T,t}) \xrightarrow{p} 0.$$

Then, under Condition 1,

$$\int \exp[\sum v_t' \delta_t + \sum \delta_t' \tilde{w}_{T,t} u + \tfrac{1}{2} \sum \delta_t' w_{T,t} \delta_t] dQ(\delta) = \exp[u' \Delta_T u] (1 + D)$$

where $\Delta_T \xrightarrow{p} 0$ and $D \xrightarrow{p} 0$ do not depend on u.

Proof. Recall that $\delta_T \overset{d}{=} \left(T^{-1} M_e F \otimes \Omega^{1/2}\right) \varepsilon$, where $\varepsilon \sim N(0, I_{kT})$. Let

$$\begin{aligned} v_T' &= \left(v_{T,1}' \Omega^{1/2}, ..., v_{T,T}' \Omega^{1/2}\right)', \\ S_T &= \left(I_T \otimes \Omega^{1/2}\right) \begin{pmatrix} w_{T,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_{T,T} \end{pmatrix} \left(I_T \otimes \Omega^{1/2}\right), \\ \tilde{S}_T &= \left(I_T \otimes \Omega^{1/2}\right) \begin{pmatrix} \tilde{w}_{T,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{w}_{T,T} \end{pmatrix}, \end{aligned}$$

and let $\mu_T := \left(I_{Tk} - 2T^{-2} A_T\right)^{-1} \left(T^{-1} M_e F \otimes I_k\right) \left(v + \tilde{S}_T (e_T \otimes u)\right)$, $A_T := (F' M_e \otimes I_k) S_T (M_e F \otimes I_k)$. Let $(\lambda_{i,t})_{i \leq k, t \leq T}$ denote the eigenvalues of $-2T^{-2} A_T$. It will be shown that and $\sup\limits_{i,t \leq T} |\lambda_{i,t}| \leq$

$\sqrt{\sum \lambda_{i,t}^2} = o_p(1)$, so that the matrix $\left(I_{Tk} - 2T^{-2}A_T\right)$ is invertible with probability approaching one. Then

$$
\begin{aligned}
& E_\delta \exp\left(\sum v_t'\delta_t + \sum \delta_t'\tilde{w}_{T,t}u + \tfrac{1}{2}\sum \delta_t' w_{T,t}\delta_t\right) \\
= \; & E_\delta \exp\left(\delta'\left(I_T \otimes \Omega^{-1/2}\right)\left(v + \tilde{S}_T\left(e_T \otimes u\right)\right) + \tfrac{1}{2}\delta'\left(I_T \otimes \Omega^{-1/2}\right) S_T \left(I_T \otimes \Omega^{-1/2}\right)\delta\right) \\
= \; & E_\varepsilon \exp\left(\varepsilon'\left(T^{-1}F'M_e \otimes I_k\right)\left(v + \tilde{S}_T\left(e_T \otimes u\right)\right) + \tfrac{1}{2}\varepsilon'\left(T^{-1}F'M_e \otimes I_k\right) S_T \left(T^{-1}M_e F \otimes I_k\right)\varepsilon\right) \\
= \; & E_\varepsilon \exp\left(\varepsilon'\left(T^{-1}F'M_e \otimes I_k\right)\left(v + \tilde{S}_T\left(e_T \otimes u\right)\right) + T^{-2}\tfrac{1}{2}\varepsilon' A_T \varepsilon\right) \\
= \; & \int d\varepsilon (2\pi)^{-kT/2} \exp\left(-\frac{1}{2}\varepsilon'\varepsilon + \varepsilon'\left(T^{-1}F'M_e \otimes I_k\right)\left(v + \tilde{S}_T\left(e_T \otimes u\right)\right) + T^{-2}\frac{1}{2}\varepsilon' A_T \varepsilon\right) \\
= \; & \int d\varepsilon (2\pi)^{-kT/2} \exp\left(-\frac{1}{2}\left(\varepsilon - \mu\right)'\left(I_{Tk} - T^{-2}2A_T\right)\left(\varepsilon - \mu\right) + \frac{1}{2}\mu'\left(I_{Tk} - 2T^{-2}A_T\right)\mu\right) \\
= \; & \exp\left(\frac{1}{2}\mu'\left(I_{Tk} - 2T^{-2}A_T\right)\mu\right)\left|I_{Tk} - 2T^{-2}A_T\right|^{1/2}.
\end{aligned}
$$

Note that, by taking a second order Taylor expansion in the neighborhood of 1 for each (i,t),

$$
\begin{aligned}
\ln\left|I_{kT} - 2T^{-2}A_T\right| &= \sum\nolimits_{i,t} \ln\left(1 + \lambda_{i,t}\right) \\
&= \sum\nolimits_{i,t} \lambda_{i,t} - \frac{1}{2}\sum\nolimits_{i,t} \lambda_{i,t}^2 \frac{1}{\left(1 + m_{i,t}\right)^2},
\end{aligned}
$$

where $m_{i,t} \in [1 - \lambda_{i,t}, 1 + \lambda_{i,t}]$. Now, Lemma (4) shows that

$$
\begin{aligned}
-\frac{1}{2}\sum\nolimits_{i,t} \lambda_{i,t} &= T^{-2} tr\left(\left(F'M_e \otimes I_k\right) S_T \left(M_e F \otimes I_k\right)\right) \\
&= T^{-2}\sum\nolimits_{i \leq k} tr\left(F'M_e S_T^i M_e F\right) \\
&= o_p(1),
\end{aligned}
$$

(where the diagonal $T \times T$ matrix S_T^i has the elements S_T corresponding to the ith parameter on the diagonal) and

$$
\begin{aligned}
-\frac{1}{2}\sum\nolimits_{i,t} \lambda_{i,t}^2 &= T^{-4} tr\left(\left(\left(F'M_e \otimes I_k\right) S_T \left(M_e F \otimes I_k\right)\right)^2\right) \\
&= T^{-4}\sum\nolimits_{i \leq k} tr\left(\left(F'M_e S_T^i M_e F\right)^2\right) \\
&= o_p(1).
\end{aligned}
$$

So, given that $\sup\limits_{i,t \leq T} |\lambda_{i,t}| \leq \sqrt{\sum \lambda_{i,t}^2} = o_p(1)$ and $m_{i,t} \in [1 - \lambda_{i,t}, 1 + \lambda_{i,t}]$, the Taylor expansion's remainder is bounded by $\left(\sup\limits_{i,t} \frac{1}{\left(1+m_{i,t}\right)^2}\right)\sum_{i,t} \lambda_{i,t}^2 = O_p(1)\, o_p(1)$ with probability approaching one. In addition, $\left|I_{kT} - 2T^{-2}A_T\right|$ is independent of u, and we may therefore conclude that $\ln\left|I_{kT} - 2T^{-2}A_T\right| = o_p(1)$ uniformly in u, i.e.

$$
\left|I_{kT} - 2T^{-2}A_T\right| = 1 + o_p(1).
$$

Now, we have

$$\left(T^{-1}M_eF\otimes I_k\right)v=\left(I_T\otimes\Omega^{1/2}\right)T^{-1}\left((v_1-\bar{v})'\cdots(\textstyle\sum_{t=1}^{T}v_t-\bar{v})'\right)'$$

and

$$\left(T^{-1}M_eF\otimes I_k\right)\tilde{S}_T\left(e_T\otimes u\right)=\left(I_T\otimes\Omega^{1/2}\right)T^{-1}\left(u'(\tilde{w}_1-\bar{w})'\cdots u'(\textstyle\sum_{t=1}^{T}\tilde{w}_t-\bar{w})'\right)'$$

where $\bar{v}=T^{-1}\sum_{t=0}^{T-1}v_{t+1}\left(T-t\right)$ and $\bar{w}=T^{-1}\sum_{t=0}^{T-1}\tilde{w}_{t+1}\left(T-t\right)$. The quadratic forms in those processes are bounded:

$$\begin{aligned}&v'\left(T^{-1}F'M_e\otimes I_k\right)\left(I_{Tk}-2T^{-2}A_T\right)^{-1}\left(T^{-1}M_eF\otimes I_k\right)v\\ \leq\ &\sup_{i,t}\left|(1+\lambda_{i,t})^{-1}\right|v'\left(T^{-1}F'M_e\otimes\Omega\right)\left(T^{-1}M_eF\otimes I_k\right)v.\end{aligned}$$

As was argued above, $\sup\limits_{i,t}\left|(1+\lambda_{i,t})^{-1}\right|=O_p\left(1\right)$. Let $\bar{\lambda}\left(\Omega\right)$ be tha largest eigenvalue of Ω. Since the process $\lambda\mapsto T^{-1/2}\left(\sum_{t=1}^{[\lambda T]}v_t-\bar{v}\right)$ converges to zero by assumption,

$$\begin{aligned}v'\left(T^{-1}F'M_e\otimes I_k\right)\left(T^{-1}M_eF\otimes I_k\right)v&=T^{-1}\sum_t\left(T^{-1/2}\left(\textstyle\sum_{l=1}^{t}v_l-\bar{v}\right)\right)'\Omega\left(T^{-1/2}\left(\textstyle\sum_{l=1}^{t}v_l-\bar{v}\right)\right)\\&\leq\bar{\lambda}\left(\Omega\right)T^{-1}\sum_t T^{-1}\left(\textstyle\sum_{l=1}^{t}v_l-\bar{v}\right)'\left(\textstyle\sum_{l=1}^{t}v_l-\bar{v}\right)\\&\leq\bar{\lambda}\left(\Omega\right)T^{-1}\sup_t\left(\textstyle\sum_{l=1}^{t}v_l-\bar{v}\right)'\left(\textstyle\sum_{l=1}^{t}v_l-\bar{v}\right)\\&=o_p\left(1\right).\end{aligned}$$

By the same token,

$$\begin{aligned}&\left(e_T'\otimes u'\right)\tilde{S}_T\left(T^{-1}F'M_e\otimes I_k\right)\left(T^{-1}M_eF\otimes I_k\right)\tilde{S}_T\left(e_T\otimes u\right)\\=\ &T^{-1}\sum_t\left(T^{-1/2}\left(\textstyle\sum_{l=1}^{t}u'(\tilde{w}_1-\bar{w})\right)\right)\Omega\left(T^{-1/2}\left(\textstyle\sum_{l=1}^{t}(\tilde{w}_1-\bar{w})u\right)\right)\\\leq\ &u'\left(\bar{\lambda}\left(\Omega\right)T^{-1}\sum_t T^{-1/2}\left(\textstyle\sum_{l=1}^{t}(\tilde{w}_1-\bar{w})\right)T^{-1/2}\left(\textstyle\sum_{l=1}^{t}(\tilde{w}_1-\bar{w})\right)\right)u\\=\ &u'\Delta_Tu\end{aligned}$$

where $\Delta_T\xrightarrow{p}0$ and is independent of u.

Hence,

$$E_\delta\exp\left(\sum v_t'\delta_t+\sum\delta_t'\tilde{w}_{T,t}u+\tfrac{1}{2}\sum\delta_t'w_{T,t}\delta_t\right)=\left(1+o_p\left(1\right)\right)\exp u'\Delta_Tu.$$

■

Lemma 4 *Under the assumptions of Lemma (3),*

$$\begin{aligned}\textstyle\sum_{i\leq k}\mathrm{tr}\left(F'M_eS_T^iM_eF\right)&=o_p\left(T^2\right)\\\textstyle\sum_{i\leq k}\mathrm{tr}\left(\left(F'M_eS_T^iM_eF\right)^2\right)&=o_p\left(T^4\right)\end{aligned}$$

Proof. For notational simplicity, we set $\Omega = I_k$, assume that $k = 1$ and ignore indexing the parameters by i.

Expand $tr\left(F'M_eS_TM_eF\right)$ into

$$\begin{aligned} & tr\left(F'S_TF\right) - 2T^{-1}tr\left(F'S_Tee'F\right) + T^{-2}tr\left(F'ee'S_Tee'F\right) \\ = \;& tr\left(F'S_TF\right) - 2T^{-1}e'FF'S_Te + T^{-2}\left(e'S_Te\right)\left(e'FF'e\right) \end{aligned}$$

Let $\sigma_t := \Omega^{1/2}\left(\sum_{l=1}^{t} w_{T,l}\right)\Omega^{1/2}$, $\tilde{\sigma}_1 := \sigma_T$, and, for $t = 2, ..., T$, $\tilde{\sigma}_t := \sigma_T - \sigma_{t-1}$, and remark that $\sup_\lambda \sigma_{[\lambda T]/T} = o_p\left(T\right)$. With

$$\Upsilon := F'S_TF = \begin{pmatrix} \tilde{\sigma}_1 & \tilde{\sigma}_2 & \cdots & \tilde{\sigma}_T \\ \tilde{\sigma}_2 & \tilde{\sigma}_2 & \cdots & \tilde{\sigma}_T \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\sigma}_T & \tilde{\sigma}_T & \cdots & \tilde{\sigma}_T \end{pmatrix}$$

so that $\left|tr\left(F'S_TF\right)\right| \leq T\sup_{t\leq T}\left|\tilde{\sigma}_t\right| = o_p\left(T^2\right)$. Moreover,

$$\left|e'S_TFF'e\right| = \left|\sum\nolimits_t \left(T-t+1\right)\tilde{\sigma}_t\right| \leq O\left(T^2\right)\sup_{t\leq T}\left|\tilde{\sigma}_t\right| = o_p\left(T^3\right),$$

$e'S_Te = \sigma_T = o_p\left(T\right)$, and $e'FF'e = \sum t^2 = O_p\left(T^3\right)$ allow to conclude that $tr\left(F'M_eS_TM_eF\right) = o_p\left(T^2\right) + T^{-1}o_p\left(T^3\right) + T^{-2}o_p\left(T\right)O_p\left(T^3\right) = o_p\left(T^2\right)$.

Expand $tr\left(\left(F'M_eS_TM_eF\right)^2\right)$ into

$$\begin{aligned} & tr\left(\Upsilon^2\right) - 4T^{-1}\left(e'FF'S_TFF'W_Te\right) + 2T^{-2}\left(e'FF'e\right)\left(e'S_TFF'S_Te\right) \\ & +2T^{-2}\left(e'S_Te\right)\left(e'FF'S_TFF'e\right) + 2T^{-2}\left(e'S_TFF'e\right)^2 \\ & -4T^{-3}\left(e'S_Te\right)\left(e'S_TFF'e\right)\left(e'FF'e\right) + T^{-4}\left(e'FF'e\right)^2\left(e'S_Te\right)^2, \end{aligned}$$

the sum of seven terms, a_1 to a_7, the last three of which are $o_p\left(T^4\right)$ from the previous steps.

For a_1,

$$\Upsilon^2 = \begin{pmatrix} \sum_{t=1}^{T}\tilde{\sigma}_t^2 & & \cdots & & \\ & \ddots & & & \\ \vdots & & j\sigma_j^2 + \sum_{t=j+1}^{T}\tilde{\sigma}_t^2 & & \vdots \\ & & & \ddots & \\ & & \cdots & & T\tilde{\sigma}_T \end{pmatrix}$$

so that

$$\begin{aligned} a_1 &= tr\left(\Upsilon^2\right) = \sum\nolimits_{t=1}^{T}\left(2t-1\right)\tilde{\sigma}_t^2 \\ &\leq \left(\sum\nolimits_{t=1}^{T} 2t - T\right)\left(\sup_{t\leq T}\left|\tilde{\sigma}_t\right|\right)^2 = O_p(T^2)o_p\left(T^2\right) \\ &= o_p\left(T^4\right). \end{aligned}$$

From $e'S_T F F' S_T e = \sum_t \sigma_t^2 = o_p\left(T^3\right)$ we can conclude that $a_3 = o_p\left(T^4\right)$. Finally,

$$\begin{aligned}
&\left|e'FF'S_TFF'e\right| \\
&\leq \quad e'F \begin{pmatrix} \sum_{t=1}^{T} |\tilde{\sigma}_t| (T-t+1) \\ \vdots \\ |\tilde{\sigma}_j| \sum_{t=1}^{j} (T-t+1) + \sum_{t=j+1}^{T} |\tilde{\sigma}_t| (T-t+1) \\ \vdots \\ |\tilde{\sigma}_T| T(T+1)/2 \end{pmatrix} \\
&= \quad \left(\sup_t |\tilde{\sigma}_t|\right) e'F \begin{pmatrix} O_p\left(T^2\right) \\ \vdots \\ O_p\left(T^2\right) \end{pmatrix} \\
&= \quad o_p(T)\, O_p\left(T^4\right) = o_p\left(T^5\right)
\end{aligned}$$

gives $a_2 = o_p\left(T^4\right)$ and $a_3 = o_p\left(T^4\right)$. ■

Lemma 5 *Let $p \geq 1$, and $\widetilde{LR}_T(\delta) = \exp\left(\sum s_t(\theta_0)' \delta_t - \frac{1}{2}\sum \delta_t' H \delta_t\right)$. Then, under θ_0 stable,*

$$E_\delta\left(\widetilde{LR}_T^p(\delta) - \widehat{LR}_T^p(0,\delta)\right) = o_p(1).$$

Proof. Recall that $\hat{u} = T^{1/2}\left(\hat{\theta} - \theta_0\right) = \tilde{H}^{-1} T^{-1/2} \sum s_t(\theta_0)$. Because of the normalization constraint on δ, $\sum_t \left(T^{-1}\sum_{l=1}^{T} s_l'(\theta_0)\right)\delta_t = 0$, and so

$$\begin{aligned}
&\left|E_\delta\left(\widetilde{LR}_T^p(\delta) - \widehat{LR}_T^p(0,\delta)\right)\right|^2 \leq E_\delta \exp\left(2p\sum s_t(\theta_0)'\delta_t\right) \\
&\times E_\delta\left[1 - \exp\left(2p\sum\left(s_t(\hat{\theta}) - s_t(\theta_0) + T^{-1}\sum_{l=1}^{T} s_l(\theta_0)\right)'\delta_t - p\sum \delta_t'(\hat{H} - H)\delta_t\right)\right]^2 \\
= \quad & E_\delta \exp\left(2p\sum s_t(\theta_0)'\delta_t\right) \\
&\times E_\delta\left[1 - \exp\left(2pT^{-1/2}\hat{u}'\sum\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)\delta_t - p\sum\delta_t'(\hat{H} - H)\delta_t\right)\right]^2.
\end{aligned}$$

Now, an application of Lemma (3) with $w_{T,t} = -p(\hat{H} - H)$, $v_{T,t} = 2pT^{-1/2}\hat{u}'\sum\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)$, and $u = 0$ gives $E_\delta \exp(2pT^{-1/2}\hat{u}'\sum\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)\delta_t - p\sum\delta_t'(\hat{H} - H)\delta_t) = 1 + o_p(1)$ and similarly for $E_\delta \exp(4pT^{-1/2}\hat{u}'\sum\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)\delta_t - 2p\sum\delta_t'(\hat{H} - H)\delta_t)$. Hence

$$E_\delta\left[1 - \exp\left(2p\hat{u}'\sum\left(h_t(\tilde{\theta}^2) + \tilde{H}\right)'\delta_t - p\sum\delta_t'(\hat{H} - H)\delta_t\right)\right]^2 = o_p(1).$$

An application of Lemma (1) gives $E_\delta \exp\left(2p\sum s_t(\theta_0)'\delta_t\right) = O_p(1)$ since an FCLT applies to the score sequence under θ_0. The result now follows. ■

Lemma 6 *For all $\theta_0 \in \Theta$, under θ_0 stable, we have (a) For any $p \geq 1$, $E_\delta \widetilde{LR}_T^p(\delta) = O_p(1)$; (b) For any $p \geq 1$, $E_\delta \widehat{LR}_T^p(0,\delta) = O_p(1)$; (c) For any $p \geq 1$, $\left(E_\delta \widetilde{LR}_T^p(\delta)\right)^{-1} = O_p(1)$.*

Proof. Part (b) follows from part (a) and Lemma (5).

For part (a), Lemma (7) shows that $E_\delta \widetilde{LR}_T(\delta)$ takes on a particular form for which arguments in all points (but for notation) similar to those of the proof of Lemma 2 in Elliott and Müller (2003), show that $E_\delta \widetilde{LR}_T(\delta) \xrightarrow{d} \overline{V}(\theta_0) = \prod_{i=1}^{k} \left(\frac{2a_i e^{-a_i}}{1-e^{-2a_i}}\right)^{\frac{1}{2}} \exp(\chi)$, where $\chi = O_p(1)$ is the (weak) limit of $-\frac{1}{2}\sum_{i=1}^{k} s_{0,i}^{*\prime}(G_{a_i} - M_e) s_{0,i}^{*}$. The same type of argument applies for $p > 1$, to show that $E_\delta \widetilde{LR}_T^p(\delta) \xrightarrow{d} \overline{V}_p(\theta_0) := C_P^{\frac{1}{2}} \exp(\chi_p)$ where $\chi_p = O_p(1)$.

For part (c), given that distribution of $\overline{V}(\theta_0)$ is absolutely continuous (see Elliott and Müller (2003)), one can apply the continuous mapping theorem to show that has $\left(E_\delta \widetilde{LR}_T^p(\delta)\right)^{-1} \xrightarrow{d} \overline{V}_p(\theta_0)^{-1}$. ■

Lemma 7 *Let* $s_0^* := \left(M_e \otimes P^{*\prime} H^{-1/2}\right) s_0$ *be the vector of the demeaned normalized scores and* $s_i^*\left(\theta^0\right)$ *its subvector corresponding the ith parameter. Then*

$$E_\delta \widetilde{LR}_T(\delta) = \prod_{i=1}^{k} \left(\frac{2a_i e^{-a_i}}{1-e^{-2a_i}}\right)^{\frac{1}{2}} \exp\left(g_J(T^{-1/2}\sum_{t=1}^{[T\cdot]} s_{i,t}^*\left(\theta^0\right))\right) + o_p(1),$$

where

$$\begin{aligned} g_J\left(T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t\right) &= -aJ_u(1)^2 - a^2\int J_u^2 - \frac{2a}{1-e^{-2a}}\left(e^{-a}J_u(1) + a\int e^{-as}J_u ds\right)^2 \\ &\quad + \left(J_u(1) + a\int J_u\right)^2 \end{aligned}$$

and

$$J_u(s) := T^{-1/2}\sum_{t=1}^{[Ts]} u_t - a\int_0^s e^{-a(s-\lambda)}\left(T^{-1/2}\sum_{t=1}^{[T\lambda]} u_t\right) d\lambda.$$

Proof. For each T, an application of Lemma (8) with $C = H$ and $\upsilon = s_0$ gives

$$\begin{aligned} E_\delta \widetilde{LR}_T(\delta) &= E_\delta \exp\left(s_0'\delta - \frac{1}{2}\delta'(I_T \otimes H)\delta\right) \\ &= \prod_{i=1}^{k}\left(\frac{1-r_{a_i}^{2T}}{T(1-r_{a_i}^2)r_{a_i}^{T-1}}\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} s_{0,i}^{*\prime}(G_{a_i} - M_e) s_{0,i}^*\right). \end{aligned}$$

Now, by assumption (FCLT), a functional central limit theorem applies to the sequence of scores $\left(s_t(\theta^0)\right)_{t\geq 1}$. Hence, the k processes $\left(\left(I \otimes P^{*\prime}H^{-1/2}\right) s_0\right)_{i,t}$ satisfy the conditions of Lemma 6 in Elliott and Müller (2003). Moreover, $\left(\frac{1-r_{a_i}^{2T}}{T(1-r_{a_i}^2)r_{a_i}^{T-1}}\right)^{-\frac{1}{2}} \to \left(\frac{2a_i e^{-a_i}}{1-e^{-2a_i}}\right)^{\frac{1}{2}}$, and the result follows from Lemma 6 in Elliott and Müller (2003). ■

Lemma 8 *Let* C *be a* $k \times k$ *positive definite, symmetric matrix,* $\upsilon = (\upsilon_1', ..., \upsilon_T')'$ *a* $Tk \times 1$ *vector, and* $P_C^* diag(c_1 \cdots c_k) P_C^{*\prime}$ *be the spectral decomposition of* $C^{1/2}\Omega C^{1/2}$*; let* $\tilde{\upsilon} = \left(M_e \otimes P_C^{*\prime} C^{-1/2}\right)$ *and*

$\tilde{\upsilon}_i$ its $T \times 1$ sub-vector corresponding the ith component of υ_t; let $K(C) = \prod_{i=1}^{k} \left(\frac{1-r_{c_i}^{2T}}{T(1-r_{c_i}^2) r_{c_i}^{T-1}} \right)^{-1/2}$ *and*

$$\Sigma(C) = \left(M_e \otimes C^{-1/2} P_C^* \right) \left(\sum_{i=1}^{k} (M_e - G_{c_i}) \otimes (\iota_{k,i} \iota_{k,i}') \right) \left(M_e \otimes P_C^{*\prime} C^{-1/2} \right).$$

Then, *(i) for any $h \in L_1(Q)$,*

$$\begin{aligned} & \int h(\delta) \exp\left(\sum \upsilon_t' \delta_t - \frac{1}{2} \sum \delta_t' C \delta_t \right) dQ(\delta) \\ = & \prod_{i=1}^{k} \left(\frac{1 - r_{c_i}^{2T}}{T(1 - r_{c_i}^2) r_{c_i}^{T-1}} \right)^{-1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{k} \tilde{\upsilon}_i' (G_{c_i} - M_e) \tilde{\upsilon}_i \right) \\ & \times \int h\left((\Sigma(C)^{1/2})^+ \delta + \Sigma(C) \upsilon \right) d\Phi_{Tk}(\delta) \\ = & K(C) \exp\left(\frac{1}{2} \upsilon' \Sigma(C) \upsilon \right) \int h\left((\Sigma(C)^{1/2})^+ (\delta + \Sigma(C) \upsilon) \right) d\Phi_{Tk}(\delta) \\ = & K(C) \exp\left(\frac{1}{2} \upsilon' \Sigma(C) \upsilon \right) \int h(\delta) \, dN\left(\Sigma(C) \upsilon, \Sigma(C)^+ \right) \end{aligned}$$

and *(ii) the density of $N(\Sigma \hat{s}, \Sigma^+)$ can be written as*

$$\frac{\widehat{LR}_T(0,\delta) dQ(\delta)}{\int \widehat{LR}_T(0,\delta) dQ(\delta)}.$$

Proof. (i) Recall that $\underset{T \times (T-1)}{B_e}$ satisfies $B_e B_e' = M_e$ and $B_e' B_e = I_{T-1}$. Define $K_\Omega := T^{-2} B_e' F F' B_e \otimes \Omega$, $\Lambda := diag(c_1 \cdots c_k)$, $K_\Lambda := T^{-2} B_e' F F' B_e \otimes \Lambda$, $\Xi^{-1} := K_\Omega^{-1} + (I_{T-1} \otimes C)$ and $\mu(\upsilon) := \Xi(B_e' \otimes I_k) \upsilon$, and let $\underset{k(T-1) \times 1}{\delta^*} \sim N(0, K_\Omega)$. From Lemma (1) in Elliott and Müller (2003),

$$(B_e \otimes I_k) \left(K_\Lambda^{-1} + (I_{T-1} \otimes I_k) \right)^{-1} (B_e' \otimes I_k) = \sum_{i=1}^{k} (M_e - G_{c_i}) \otimes (\iota_{k,i} \iota_{k,i}')$$

so that

$$\begin{aligned} (B_e \otimes I_k) \Xi (B_e' \otimes I_k) &= (B_e \otimes I_k) \left(I_{T-1} \otimes C^{-1/2} P_C^* \right) (B_e' \otimes I_k) (B_e \otimes I_k) \left(K_\Lambda^{-1} + (I_{T-1} \otimes I_k) \right)^{-1} \\ & \quad \times (B_e' \otimes I_k) (B_e \otimes I_k) \left(I_{T-1} \otimes P_C^{*\prime} C^{-1/2} \right) (B_e' \otimes I_k) \\ &= \left(M_e \otimes C^{-1/2} P_C^* \right) \left(\sum_{i=1}^{k} (M_e - G_{c_i}) \otimes (\iota_{k,i} \iota_{k,i}') \right) \left(M_e \otimes P_C^{*\prime} C^{-1/2} \right) \\ &= \left(I_T \otimes C^{-1/2} P_C^* \right) \left(\sum_{i=1}^{k} (M_e - G_{c_i}) \otimes (\iota_{k,i} \iota_{k,i}') \right) \left(I_T \otimes P_C^{*\prime} C^{-1/2} \right) \\ &= \Sigma(C), \end{aligned}$$

Moreover, from Lemma 4 (iii) in Elliott and Müller (2003),

$$\begin{aligned}|K_\Omega|^{-1/2}\left|\Xi^{-1}\right|^{-1/2} &= \left|K_\Omega\left(K_\Omega^{-1}+I\right)\right|^{-1/2}\\ &= |K_\Omega+I|^{-1/2}\\ &= \left|B_e' H_a B_e\right|^{-1/2}\\ &= \left(\frac{1-r_{c_i}^{2T}}{T(1-r_{c_i}^2)r_{c_i}^{T-1}}\right)^{-1/2}.\end{aligned}$$

Also, from Condition (1), $(B_e\otimes I_k)\,\delta^* \overset{d}{=} \delta$ so that

$$\begin{aligned}&\int h(\delta)\exp\left(\sum \upsilon_t'\delta_t-\frac{1}{2}\sum\delta_t' C\delta_t\right)dQ(\delta)\\ =\;& E_\delta h(\delta)\exp\left(\upsilon'\delta-\frac{1}{2}\delta'\left(I_T\otimes C\right)\delta\right)\\ =\;& E_{\delta^*} h\left(\left(B_e\otimes I_k\right)\delta^*\right)\exp\left(\upsilon'\left(B_e\otimes I_k\right)\delta^*-\frac{1}{2}\delta^{*\prime}\left(B_e'\otimes I_k\right)\left(I_T\otimes C\right)\left(B_e\otimes I_k\right)\delta^*\right)\\ =\;& \int d\delta^*\,(2\pi)^{-\frac{T-1}{2}}\,|K_\Omega|^{-1/2}\,h\left(\left(B_e\otimes I_k\right)\delta^*\right)\exp\left(-\frac{1}{2}\delta^{*\prime}K_\Omega^{-1}\delta^*\right)\\ &\times\exp\left(\upsilon'\left(B_e\otimes I_k\right)\delta^*-\frac{1}{2}\delta^{*\prime}\left(I_{T-1}\otimes C\right)\delta^*\right)\\ =\;& |K_\Omega|^{-1/2}\left|\Xi^{-1}\right|^{-1/2}\exp\left(\frac{1}{2}\upsilon'\Sigma(C)\,\upsilon\right)\\ &\times\int d\delta^*\,(2\pi)^{-\frac{T-1}{2}}\,|\Xi|^{-1/2}\,h\left(\left(B_e\otimes I_k\right)\delta^*\right)\exp\left(-\frac{1}{2}(\delta^*-\mu(\upsilon))'\Xi^{-1}(\delta^*-\mu(\upsilon))\right)\\ =\;& K(C)\exp\left(\frac{1}{2}\upsilon'\Sigma(C)\,\upsilon\right)\int h\left(\left(B_e\otimes I_k\right)\delta\right)dN(\mu(\upsilon),\Xi)\\ =\;& K(C)\exp\left(\frac{1}{2}\upsilon'\Sigma(C)\,\upsilon\right)\int h(\delta)\,dN(\Sigma(C)\,\upsilon,\Sigma(C))\\ =\;& K(C)\exp\left(\frac{1}{2}\upsilon'\Sigma(C)\,\upsilon\right)\int h\left((\Sigma(C)^{1/2})^+\left(\delta-\Sigma(C)\,\upsilon\right)\right)dN(0,I_{Tk})\end{aligned}$$

(ii) From part (i), with $h\equiv 1$, with $\upsilon=\hat{s}$, $C=\hat{H}$, $\Sigma=\Sigma(C)$, and $\widehat{LR}_T(0,\delta)=\exp\left(\hat{s}'\delta-\frac{1}{2}\delta'\hat{H}\delta\right)$ gives $\int\widehat{LR}_T(0,\delta)dQ(\delta)=K(\hat{H})\exp\left(\frac{1}{2}\hat{s}'\Sigma\hat{s}\right)$ and $\widehat{LR}_T(0,\delta)dQ(\delta)=K(\hat{H})\exp\left(\frac{1}{2}\hat{s}'\Sigma\hat{s}\right)dN(\Sigma\hat{s},\Sigma^+)$.
■

Lemma 9 *For all $\theta_0\in\not\leq$, (a) The densities $f_1^*(\theta_0):=\int f_T(\theta_0,\delta)dQ(\delta)$ and $f_0^*(\theta_0):=f_T(\theta_0,0)$ are contiguous; (b) The densities $\tilde{f}(\theta)=\kappa^{-1}\int_{\mathcal{V}_T(\theta)}f_T(\theta,\delta)dQ(\delta)$ and $f_0^*(\theta_0):=f_T(\theta_0,0)$ are contiguous.*

Proof. Part (a). By Lemma (6.4) of Vaart (1998), it suffices to show that $f_1^*(\theta_0)/f_0^*(\theta_0)$ converges in distribution under $f_0^*(\theta_0)$ to a random variable $V(\theta_0)$ such that $EV(\theta_0)=1$.

Now, $E_\delta\left(1-\mathcal{S}_T\right)\widehat{LR}_T\left(0,\delta\right)=o_p\left(1\right)$ since $\left(E_\delta\left(1-\mathcal{S}_T\right)\widehat{LR}_T\left(0,\delta\right)\right)^2\leq E_\delta\left(1-\mathcal{S}_T\right)E_\delta\widehat{LR}_T^2\left(0,\delta\right)=o_p\left(1\right)O_p\left(1\right)$ by Lemma (6), and, by Markov's inequality,

$$
\begin{aligned}
P\left(E_\delta\left(1-\mathcal{S}_T\right)LR_T\left(0,\delta\right)>\eta\right) &\leq \eta^{-1}E_\delta\left(1-\mathcal{S}_T\right)\left(\int f_T(\theta_0,\delta)d\mu_T\right) \\
&= \eta^{-1}E_\delta\left(1-\mathcal{S}_T\right)=o\left(1\right).
\end{aligned}
$$

It follows that $\left|E_\delta\left(\widehat{LR}_T\left(0,\delta\right)-LR_T\left(0,\delta\right)\right)\right|^2\leq E_\delta\mathcal{S}_T\left|\widehat{LR}_T\left(0,\delta\right)-LR_T\left(0,\delta\right)\right|^2+o_p\left(1\right)$. Note that when $u=0$, $LR_T(0,\delta)=\exp\left(\sum s_t(\hat{\theta})'\delta_t+T^{-1/2}\hat{u}'\sum h_t(\tilde{\theta}^2)\delta_t+\frac{1}{2}\sum\delta_t'h_t(\tilde{\theta}_t)\delta_t\right)$

$$
\begin{aligned}
\varsigma_T &= T^{-1/2}\hat{u}'\sum h_t(\tilde{\theta}^2)\delta_t+\tfrac{1}{2}\sum\delta_t'\left(h_t(\tilde{\theta}_t)+\hat{H}\right)\delta_t \\
LR_T(0,\delta) &= \widehat{LR}_T(0,\delta)\exp\varsigma_T \\
&= \exp\left(\sum s_t(\hat{\theta})'\delta_t-\frac{1}{2}\sum\delta_t'\hat{H}\delta_t+\varsigma_T\right).
\end{aligned}
$$

The same arguments as in the proof of part (c) of Theorem (1) can be applied to show that $E_\delta\mathcal{S}_T\left|\widehat{LR}_T\left(0,\delta\right)-LR_T\left(0,\delta\right)\right|^2=o_p\left(1\right)$. Hence, $E_\delta\widehat{LR}_T\left(0,\delta\right)-E_\delta LR_T\left(0,\delta\right)\xrightarrow{p}0$. Applying Lemma (5) then gives that $f_1^*\left(\theta_0\right)/f_0^*\left(\theta_0\right)$ and $E_\delta\widetilde{LR}_T\left(\delta\right)$ share the same limit. Lemma (7) together with Lemma 2 in Elliott and Müller (2003), shows that $E_\delta\widetilde{LR}_T\left(\delta\right)\xrightarrow{d}V\left(\theta_0\right)$ with $EV\left(\theta_0\right)=1$, which completes the proof of part (a).

Part (b). Note that $E_\delta\left(1-\mathcal{S}_T\right)\mathcal{V}_T(\theta)\widehat{LR}_T\left(0,\delta\right)=o_p\left(1\right)$ since $\left(E_\delta\left(1-\mathcal{S}_T\right)\mathcal{V}_T(\theta)\widehat{LR}_T\left(0,\delta\right)\right)^2\leq E_\delta\left(1-\mathcal{S}_T\right)E_\delta\widehat{LR}_T^2\left(0,\delta\right)$ and so

$$
\begin{aligned}
\left|E_\delta\left(\kappa^{-1}\mathcal{V}_T(\theta)LR_T\left(0,\delta\right)-\widehat{LR}_T\left(0,\delta\right)\right)\right| &\leq E_\delta\mathcal{S}_TLR_T\left(0,\delta\right)\left|\kappa^{-1}\mathcal{V}_T(\theta)-1\right| \\
&\quad +E_\delta\mathcal{S}_T\left|LR_T\left(0,\delta\right)-\widehat{LR}_T\left(0,\delta\right)\right|+o_p\left(1\right) \\
&\leq \max\left(1,\left|\kappa^{-1}-1\right|\right)E_\delta\mathcal{S}_TLR_T\left(0,\delta\right)+o_p\left(1\right) \\
&= o_p\left(1\right),
\end{aligned}
$$

which proves part (b). ■

References

Andrews, D. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

Andrews, D., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Bai, J., and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

Cooley, T., and E. Prescott (1976): "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica*, 44, 167–184.

Durbin, J., and S. Koopman (1997): "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models," *Biometrika*, 84, 669–684.

Elliott, G., and U. Müller (2003): "Optimally Testing General Breaking Processes in Linear Time Series Models," *UCSD Working Paper 2003-07.*

Engle, R., D. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Gallant, A., and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Basil Blackwell, Oxford.

Hansen, B. (2000): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115.

Harvey, A. (1981): *Time Series Models.* Philipp Allan, Oxford.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Pollard, D. (2001): "Contiguity," unpublished draft, http://www.stat.yale.edu/~pollard/Paris2001/Lectures/Contiguity.pdf.

Primiceri, G. (2004): "Time Varying Structural Vector Autoregressions and Monetary Policy," *mimeo, Princeton University.*

Schervish, M. (1995): *Theory of Statistics.* Springer, New York.

Shephard, N., and M. Pitt (1997): "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, pp. 653–667.

Sims, C. (1993): "A 9 Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators and Forecasting*, ed. by J. Stock, and M. Watson, pp. 179–214. NBER studies in business cycles 28.

Stock, J., and M. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

——— (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

Vaart, A. V. d. (1998): *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.