Efficient Estimation of the Parameter Path in Unstable Time Series Models

Ulrich Mueller and Philippe-Emmanuel Petalas

Princeton University

March 2007

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

March 2007

Abstract

The paper investigates asymptotically efficient inference in general likelihood models with time varying parameters. Parameter path estimators and tests of parameter constancy are evaluated by their weighted average risk and weighted average power, respectively. The weight function is proportional to the distribution of a Gaussian process, and focusses on local parameter instabilities that cannot be detected with certainty even in the limit. It is shown that asymptotically, the sample information about the parameter path is efficiently summarized by a Gaussian pseudo model. This approximation leads to computationally convenient formulas for efficient path estimators and test statistics, and unifies the theory of stability testing and parameter path estimation.

JEL Classification: C22, C13, C12, C11

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

---

*The authors would like to thank Mark Watson and participants at the NBER Summer Institute, the Workshop for Nonlinear and Nonstationary Models at the California Institute of Technology, the Unit Root and Cointegration Testing Conference in Faro, the Econometric Society World Congress in London, and workshops at the University of Lausanne, New York University, Rutgers University, University of Texas at Austin, FRB of Atlanta and Iowa State University. Müller gratefully acknowledges financial support from the NSF through grant SES-0518036.

1 Introduction

One of the central concerns in time series modelling is the stability of parameters through time. A large body of econometric work has developed around testing the hypothesis that parameters are time invariant: see, for instance, Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994), Bai, Lumsdaine, and Stock (1998), Vogelsang (1998), Hansen (2000), Andrews (2003) and Elliott and Müller (2006) for some more recent contributions, and Stock (1994) and Dufour and Ghysels (1996) for surveys and additional references. Empirically, there is substantial evidence of instabilities in the parameters of finance and macroeconomic models as documented in Stock and Watson (1996), Ghysels (1998), McConnell and Perez-Quiros (2000), Lettau and Ludvigson (2001), Boivin (2003), Primiceri (2005), Cogley and Sargent (2005) and Paye and Timmermann (2006), just to name a few.

Once instabilities are suspected, a natural next step is to document their form. Knowledge of the parameter path is useful for a number of purposes. First, the estimated path is an interesting descriptive tool, as it helps to understand potential sources of the instability. Second, the endpoint of the parameter path is useful for forecasting purposes (see, for instance, Chernoff and Zacks (1964), Sims (1993), Stock and Watson (1996) or Pesaran, Pettenuzzo, and Timmermann (2006)). Third, economic theory might imply certain features of parameter paths (think, for instance, of convergence models with time varying mean growth of GDP), for which one might want to test in econometric models. Finally, the time varying value of the parameter can sometimes be given a useful structural interpretation (cf. Cooley and Prescott (1976)), such as, in a regression model, the time dependent marginal effect of a certain regressor.

There are several approaches to estimating the parameter path. One develops frequentist inference for the break date in models where the parameters are known a priori to be subject to a small number of sudden shifts, such as Bai (1997), Bai and Perron (1998), and Elliott and Müller (2004). A Bayesian literature (Hamilton (1989), Chib (1998), Kim and Nelson (1999) and Sims and Zha (2006), for instance) posits a finite number of regimes for the parameter values and obtains posterior probabilities for each regime through time. Robinson (1989, 1991) and Cai (2007) develop nonparametric kernel estimators of the time varying parameter. Finally, a large frequentist and Bayesian literature estimates models under the assumption of a smooth stochastic evolution of the parameter. When the parameters enter the model linearly and disturbances are assumed Gaussian, then these models can be written in state space form and estimated by variants of Kalman filtering and smoothing, as in Cooley and Prescott (1973,

1976) and Harvey (1989). This is not possible for models with time varying parameters that affect, say, variances and covariances, and considerably more involved numerical techniques have been developed to deal with such models: see, for instance, Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1995), Durbin and Koopman (1997), Shephard and Pitt (1997), Uhlig (2001), Primiceri (2005) and Cogley and Sargent (2005) for the estimation of models with time varying second moments. In general, the estimation of time varying parameter models outside the Gaussian state space framework requires fairly complicated and model-specific numerical techniques.

This paper is closely related to this last strand. We consider a general parametric model with local time variation, in the sense that good tests would detect the instability with probability smaller than one even in the limit. We analyze estimators and tests that minimize weighted average risk and maximize weighted average power over the set of possible parameter paths, where the weighting function is proportional to the distribution function of a Gaussian process, and focusses on such local parameter variability. The main contribution is an asymptotically efficient approximation of the sample information about the parameter path. This approximation turns the problem of inference about the parameter path in the general likelihood model into the problem of inference about the parameter path in a linear Gaussian pseudo model, with the sequence of scores (evaluated at the usual maximum likelihood estimator) as the observations. Asymptotically efficient parameter path estimators and test statistics thus become straightforward to compute, and the estimation and testing problem are unified in one coherent asymptotic framework. In the special case of an underlying parametric model that is stationary for stable parameters, and a weighting that corresponds to the distribution of a Wiener process, the approximate pseudo model becomes a local level model in the sense of Harvey (1989), and optimal path estimators are obtained by an exponential smoothing of the sequence of score vectors. From a Bayesian perspective with the weighting function interpreted as the prior, our results provide an asymptotically accurate multivariate Gaussian approximation to the posterior distribution of the parameter path.

As already noted, we consider instabilities of the same magnitude as local alternatives of efficient stability tests. The asymptotic thought experiment hence leads to a limit theory where there is only limited information about the form of the instability (in contrast, say, to the set-up in Robinson (1989)). In this way, the asymptotics reflect the difficulties of not being sure about the precise form or even presence of the instability in small samples in most econometric models of interest.

Formally, in such asymptotics the magnitude of the instability decreases as the sample size increases. This does not mean that the theory developed in this paper only applies to economically insignificant instabilities. Parameter variations that are 'small' in the statistical sense of being nontrivial to detect need not be small in an economic sense. For instance, in a stylized model, a sudden shift of 1.2 percentage points in yearly GDP mean growth in the middle of a sample of 180 quarterly observations is detected less than half the time by 5% level efficient stability tests (Elliott and Müller (2004)), yet such a shift is arguably of major economic (and policy) relevance. Many instabilities that economists care about, such as those arising from Lucas-critique arguments (for instance Linde (2001)), the stability of monetary policy (for instance Bernanke and Mihov (1998)) or reduced form bivariate econometric relationships between macroeconomic variables in general (Stock and Watson (1996)) have been difficult (or at least nontrivial) to determine empirically and are hence 'small' in the statistical sense. In these instances, accurate approximations are generated by a modelling strategy in which correspondingly there is only limited statistical information about the instability asymptotically.

Our results are driven by a quadratic approximation to the log-likelihood of the general model. Such approximations of the likelihood for models with a finite dimensional parameter have a long history in statistics and econometrics and allow the substitution of a complex decision problem by a simpler one; see, for instance, LeCam (1986). Recent applications of these ideas in time series econometrics include Andrews and Ploberger (1994), Phillips and Ploberger (1996), Ploberger (2004) and Phillips and Ploberger (2006). The sample information about the parameter path is more difficult to approximate, as the path is not finite dimensional. Some numerical methods for time series models with latent variables, such as those developed by Durbin and Koopman (1997) and Shephard and Pitt (1997), employ quadratic expansions of the log-likelihood at some stage, but without rigorous justification. The recent results by Carrasco, Hu, and Ploberger (2005) on efficient tests for Markov Switching type parameter instabilities also rely on higher order expansions of the likelihood; the main difference to our results concerns the weighting function, which in their case focusses on high frequency parameter variations. Brown and Low (1996) and Nussbaum (1996) prove the asymptotic equivalence of some specific infinite dimensional decision problems with the continuous time problem of observing Gaussian White Noise with some unknown drift. These papers (essentially) establish the asymptotic equivalence of the frequentist risk function for any bounded loss function. Compared to this literature, our results are more specific, as we only show equivalence with respect to weighted average risk, where the weighting functions correspond to the distribution of a (finite mixture of) Gaussian processes. At the same time, our results are substantially more general, as they apply to a wide class of parametric time series models.

The remainder of the paper is organized as follows. The next section gives a heuristic argument for the approximation of the sample information with a linear Gaussian pseudo model, and provides the computational details for asymptotically efficient parameter path estimators and tests under a Wiener process weighting function. Section 3 contains the formal discussion of our results, and Section 4 concludes. All proofs are collected in an appendix.

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

Consider a stationary and stable time series model with known log-likelihood function of the form $\sum_{t=1}^{T} l_t(\theta)$, with parameter $\theta \in \Theta \subset \mathbb{R}^k$. The corresponding unstable model has the same likelihood with time varying parameter $\{\theta_t\}_{t=1}^T = \{\theta + \delta_t\}_{t=1}^T$. Suppose the researcher is interested in obtaining good path estimators under a weighted average risk criterion with a weighting function that is diffuse for the benchmark value θ, and posits a weighting function of a Gaussian process of magnitude $T^{-1/2}$ for the deviations $\{\delta_t\}_{t=1}^T$.

The sample information about the path $\{\theta + \delta_t\}_{t=1}^T$ is fully contained in the function $\sum l_t(\theta + \delta_t)$, where '$\sum$' denotes a sum over $t = 1, \cdots, T$. Let $\hat{\theta}$ be the maximum likelihood estimator of θ ignoring parameter instability, i.e. $\hat{\theta}$ maximizes $\sum l_t(\theta)$. Denote by $s_t(\theta) = \partial l_t(\theta)/\partial\theta$ the sequence $t = 1, \cdots, T$ of $k \times 1$ score vectors, and by $h_t(\theta) = -\partial s_t(\theta)/\partial\theta'$ the sequence of $k \times k$ Hessians. By T second order Taylor expansions

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

where $\tilde{\theta}_t$ lies on the line segment between $\theta + \delta_t$ and $\hat{\theta}$. 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 $\hat{H} = T^{-1}\sum h_t(\hat{\theta})$. Since the deviations $\{\delta_t\}_{t=1}^T$ are persistent and of order $T^{-1/2}$, and the maximum likelihood estimator $\hat{\theta}$ is a $\sqrt{T}$ consistent estimator of the benchmark value θ, the sequence $\{\theta + \delta_t - \hat{\theta}\}_{t=1}^T$ is persistent and of order $T^{-1/2}$. Also, because the stable model is assumed stationary, smooth averages of $h_t(\tilde{\theta}_t)$ are close to $\hat{H}$ in all parts of the sample, so that

$$\sum(\theta + \delta_t - \hat{\theta})'h_t(\tilde{\theta}_t)(\theta + \delta_t - \hat{\theta}) \simeq \sum(\theta + \delta_t - \hat{\theta})'\hat{H}(\theta + \delta_t - \hat{\theta}) \qquad (1)$$

and we can write

$$\begin{aligned}&\textstyle\sum(l_t(\theta+\delta_t)-l_t(\hat\theta)-\frac{1}{2}s_t(\hat\theta)'\hat H^{-1}s_t(\hat\theta))\\&\qquad\simeq-\tfrac{1}{2}\textstyle\sum(s_t(\hat\theta)-\hat H(\theta+\delta_t-\hat\theta))'\hat H^{-1}(s_t(\hat\theta)-\hat H(\theta+\delta_t-\hat\theta)).\end{aligned}\tag{2}$$

Neither $\sum l_t(\hat\theta)$ nor $\sum s_t(\hat\theta)'\hat H^{-1}s_t(\hat\theta)$ depend on $\{\theta+\delta_t\}_{t=1}^T$, so that ignoring these constants, the log-likelihood of the path $\{\theta+\delta_t\}_{t=1}^T$ is well approximated by a quadratic form. In fact, the right-hand side of (2) is recognized as the log-likelihood function of the Gaussian random variable $s_t(\hat\theta)+\hat H\hat\theta$ with mean $\theta+\delta_t$ and covariance matrix $\hat H$. The information in the sample about $\theta+\delta_t$ can therefore be approximately summarized by the pseudo model

$$s_t(\hat\theta)+\hat H\hat\theta=\hat H(\theta+\delta_t)+\nu_t,\ t=1,\cdots,T\tag{3}$$

with $\nu_t\sim i.i.d.\mathcal{N}(0,\hat H)$. For a weighting function for the benchmark value θ that is diffuse, the weighting on the mean $T^{-1}\sum\delta_t$ in (3) has no bearing on the analysis. For convenience, one might thus assume a weighting function for $\{\delta_t\}_{t=1}^T$ that corresponds to the distribution of a *demeaned* Gaussian process (so that $\sum\delta_t=0$ and δ_t is the the deviation at date t from the average parameter value θ). Under that assumption, we trivially have $\sum\delta_t\hat H(\theta-\hat\theta)=0$, and also $\sum s_t(\hat\theta)=0$ from the first order condition of the maximum likelihood estimator. Thus, the right-hand side of (2) becomes

$$-\tfrac{1}{2}\textstyle\sum(s_t(\hat\theta)-\hat H\delta_t)'\hat H^{-1}(s_t(\hat\theta)-\hat H\delta_t)-\frac{1}{2}T(\theta-\hat\theta)'\hat H(\theta-\hat\theta)$$

and the sample information about θ and $\{\delta_t\}_{t=1}^T$ is approximately independent and described by the pseudo model

$$\hat\theta=\theta+T^{-1/2}\hat H^{-1}\nu_0\tag{4}$$

$$s_t(\hat\theta)=\hat H\delta_t+\nu_t,\quad t=1,\cdots,T\tag{5}$$

with $\nu_t\sim i.i.d.\mathcal{N}(0,\hat H)$. The approximation in (4) is the standard result that in large samples, the likelihood about a parameter converges to that of a Gaussian random variable with mean $\hat\theta$ and covariance matrix $T^{-1}\hat H^{-1}$. The focus and contribution of this paper is to argue for the Gaussian 'local level' model (5) (or, equivalently, for (3)) as an asymptotically efficient summary of the sample information about the deviations $\{\delta_t\}_{t=1}^T$, at least under Gaussian weighting functions for $\{\delta_t\}_{t=1}^T$ that put almost all of their weight on deviations of the order $T^{-1/2}$. For weighting functions for $\{\delta_t\}_{t=1}^T$ that are Markovian, the asymptotically efficient path estimator under a wide range of symmetric loss functions can hence be computed by variants of the Kalman smoother. Also, asymptotically efficient tests of parameter instability in the general likelihood model can be obtained by performing optimal tests in the pseudo models. The next section formally derives a more general asymptotic equivalence statement that does not require averages of $h_t(\hat{\theta})$ to converge in probability to the same constant in all parts of the sample under some fairly weak regularity conditions on the likelihood. The generalization is useful for, say, models with a time trend, for which (1) does not hold.

We now turn to an explicit description of the optimal parameter path estimator and test statistics assuming (1) holds for a weighting function on δ_t that is a (demeaned) multivariate Gaussian random walk. This choice of weighting function (or prior in a Bayesian context) has been used extensively in econometric applications: see, for instance, Cooley and Prescott (1976), Harvey (1989), Stock and Watson (1996, 1998, 2002), Boivin (2003) Primiceri (2005) and Cogley and Sargent (2005). Without loss of generality, let the first $p \leq k$ parameters of θ, denoted β, be those whose path is to be estimated (so that the last $k - p$ elements of δ_t are zero). Denote by $\hat{s}_{\beta,t}(\hat{\theta})$ the corresponding scores, evaluated at maximum likelihood estimator $\hat{\theta}$ (whose first p elements are denoted $\hat{\beta}$) that ignores any potential instability, i.e. $\hat{s}_{\beta,t}(\hat{\theta}) = \partial l_t(\theta)/\partial\beta|_{\theta=\hat{\theta}}$, $t = 1, \cdots, T$. Let $\hat{H}_\beta = T^{-1}\sum \hat{s}_{\beta,t}(\hat{\theta})\hat{s}_{\beta,t}(\hat{\theta})'$, which is computationally convenient and asymptotically equivalent to $-T^{-1}\sum \partial^2 l_t(\theta)/\partial\beta\partial\beta'|_{\theta=\hat{\theta}}$. Under the theoretically attractive choice of the covariance matrix of the Gaussian random walk for the first p elements of $\{\delta_t\}$ to be proportional to $\hat{H}_\beta^{-1}$ (see comment 9 in Section 3 below), an asymptotically efficient path estimator may be obtained by the following algorithm:

1. Compute the sequence $x_t = \hat{H}_\beta^{-1} s_{\beta,t}(\hat{\theta})$, $t = 1, \cdots, T$.

2. Let $z_1 = x_1$, and compute

$$z_t = r_c z_{t-1} + (x_t - x_{t-1}),\ t = 2, \cdots, T$$

   where $r_c = 1 - c/T$. That is, generate an $p \times 1$ AR(1) process initialized at x_1 and innovations Δx_t.

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

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

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

5. The efficient estimator of the parameter path for β is now given by $\{\hat{\beta} + x_t - r_c \bar{z}_t\}_{t=1}^T$.

6. An asymptotically weighted average power maximizing test for parameter stability of the first p parameters of θ can be based on the statistic $\mathrm{qLL} = \sum_{t=1}^{T}(r_c\bar{z}_t - x_t)' s_{\beta,t}(\hat{\theta})$, where stability is rejected for small values.

This procedure depends on the positive parameter c, which corresponds to the signal-to-noise ratio in the smoothing problem: The smaller c, the smoother the estimated parameter path $\{\hat{\beta}+x_t-r_c\bar{z}_t\}_{t=1}^T$ becomes. One approach is to fix c at some value to obtain point optimal path estimators and tests. Elliott and Müller (2006) suggest a value of $c = 10$ for this testing problem in the context of a linear regression, and their Table 1 contains asymptotic critical values for the statistic qLL as a function of p. For the path estimation problem, a value of $c = 10$ corresponds at least roughly to the magnitude of instabilities found in macro series, cf. Stock and Watson (1998). Stock and Watson (2002) employ a fixed value of $c = 7$ in their smoothing application.

Alternatively, one might posit a weighting function for $\{\delta_t\}$ that is a mixture of n_G Gaussian random walks with signal-to-noise ratios $c \in \{c_1, \cdots, c_{n_G}\}$. Let $\{\hat{\beta}_t(c)\}_{t=1}^T$ with $\hat{\beta}_t(c) = \hat{\beta} + x_t - r_c\bar{z}_t$, $t = 1, \cdots, T$ be the path estimator as described above for a given value of c, and let $\mathrm{qLL}(c)$ be the corresponding test statistic. The asymptotically average weighted risk minimizing path estimator under this composite weighting function with truncated quadratic loss and large truncation point is then approximately given by

$$\hat{\beta}_t = \sum_{i=1}^{n_G} w_i \hat{\beta}_t(c_i), \quad t = 1, \cdots, T$$

where $w_i = \tilde{w}_i / \sum_{j=1}^{n_G} \tilde{w}_j$ and $\tilde{w}_i = \sqrt{\frac{2c_i e^{-c_i}}{1-e^{-2c_i}}} \exp(-\frac{1}{2}\,\mathrm{qLL}(c_i))$. We suggest a default choice $c \in \{0, 5, 10, \cdots, 45\}$ with $n_G = 10$, where $\tilde{w}_1 = 1$ and $\hat{\beta}_t(0) = \hat{\beta}$, $t = 1, \cdots, T$. These values for c cover the range for the magnitude of most empirically relevant instabilities.

In many applications, it will be of interest to get some sense of the accuracy of this path estimator. One such measure for the accuracy of $\hat{\beta}_t$ for a particular time period t is given by

$$V_t = T^{-1}\hat{H}_\beta^{-1} \sum_{i=1}^{n_G} w_i \kappa_t(c_i) \tag{6}$$

where $\kappa_t(c) = c(1 + e^{2c} + e^{2ct/T} + e^{2c(1-t/T)})/(2 - 2e^{2c})$ for $c > 0$ and $\kappa_t(0) = 1$. From a Bayesian perspective with the weighting function for $\{\delta_t\}$ and θ interpreted as priors, (6) is the covariance matrix of the approximate posterior distribution of β_t. This approximate posterior distribution is a mixture of multivariate normals $\mathcal{N}(\hat{\beta}_t(c_i), T^{-1}\hat{H}_\beta^{-1}\kappa_t(c_i))$ with mixing probabilities w_i. The interval $[\hat{\beta}_{t,i} - 2\sqrt{V_{t,ii}}, \hat{\beta}_{t,i} + 2\sqrt{V_{t,ii}}]$ with $\hat{\beta}_{t,i}$ the ith element of $\hat{\beta}_t$ and

$V_{t,ii}$ the i,ith element of V_t is thus approximately a 95% credible set for the ith component of β at time t (one could, of course, also determine the exact 95% credible set for the given mixture of normals posterior, with typically very similar results). This interval is not a confidence interval in the frequentist sense, but it can be justified without explicit Bayesian reasoning as a weighted average risk minimizing set estimator—see Schervish (1995), page 329.

3 Asymptotically Efficient Inference in Unstable Time Series Models

We begin by introducing some additional notation and definitions. Consider a standard parametric model for data $y_T = (y_{T,1}, \cdots, y_{T,T}) \in \mathbb{R}^{mT}$ in a sample of size T, a random vector defined on the complete probability space $(\mathcal{F}, \mathfrak{F}, P)$, with parameter $\theta \in \Theta \subset \mathbb{R}^k$ and density $\prod_{t=1}^T f_{T,t}(\theta)$ with respect to some σ-finite measure μ_T. This form of likelihood arises naturally in the 'forecasting error decomposition' of models, where $f_{T,t}(\theta)$ is the conditional likelihood of $y_{T,t}$ given $\mathfrak{F}_{T,t-1}$, where $\mathfrak{F}_{T,t} \subset \mathfrak{F}$ is the σ-field generated by $\{y_{T,s}\}_{s=1}^t$. In models with weakly exogenous components in the sense of Engle, Hendry, and Richard (1983), $f_{T,t}(\theta)$ can be decomposed into two pieces $f_{T,t}(\theta) = f_{T,t}^1(\theta) f_{T,t}^2$, where $f_{T,t}^2$ captures the contribution of the evolution of weakly exogenous components and does not depend on θ. If this is the case, only $f_{T,t}^1(\theta)$ needs to be specified. Define $l_{T,t}(\theta) = \ln f_{T,t}(\theta)$, $s_{T,t}(\theta) = \partial l_{T,t}(\theta)/\partial\theta$ and $h_{T,t}(\theta) = -\partial s_{T,t}(\theta)/\partial\theta'$. In the following definitions and conditions, we omit the dependence on T of $\mathfrak{F}_{T,t}$, $l_{T,t}$, $s_{T,t}$, $h_{T,t}$ and so forth to enhance readability. Let $[\cdot]$ indicate the largest lesser integer function, let $||\cdot||$ denote the spectral norm, let '$\otimes$' be the Kronecker product and let '$\xrightarrow{p}$' and '$\Rightarrow$' denote convergence in probability and convergence in distribution as $T \to \infty$, respectively. Measurability is understood in the Borel sense and with respect to the Euclidean topology, if not indicated otherwise.

We assume the following condition on this model with true and stable parameter θ_0.

Condition 1 *(MEAS) The functions* $f_{T,t}^1 : \mathbb{R}^{mT} \times \Theta \mapsto \mathbb{R}$ *are jointly measurable for* $t = 1, \cdots, T$.

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

(ID) There exists $\eta > 0$ *such that for all* $\epsilon > 0$ *there exists* $K(\epsilon) > 0$ *for which* $P(\sup_{||\theta-\theta_0||\geq\epsilon} T^{-1} \sum \sup_{||v||<T^{-1/2+\eta}, \theta+v\in\Theta}(l_t(\theta+v) - l_t(\theta_0)) < -K(\epsilon)) \to 1$

(LLLN) (i) For any decreasing ball of θ*, i.e.* $\mathcal{B}_T = \{\theta : ||\theta - \theta_0|| < b_T\}$ *for some sequence of real numbers* $b_T \to 0$*,* $T^{-1}\sum_{t=1}^T \sup_{\theta \in \mathcal{B}_T} ||h_t(\theta) - h_t(\theta_0)|| \xrightarrow{p} 0$*, (ii)* $T^{-1}\sum_{t=1}^T ||h_t(\theta_0)|| = O_p(1)$ *and (iii)* $\sup_{\lambda\in[0,1]} \left\| T^{-1}\sum_{t=1}^{[\lambda T]} h_t(\theta_0) - \int_0^\lambda \Gamma(l)dl \right\| \xrightarrow{p} 0$ *for some nonstochastic matrix function* Γ *(possibly indexed by* θ_0*), with* $\Gamma(\lambda)$ *positive definite for all* $\lambda \in [0,1]$*.*

(MDA) $\{s_t(\theta_0), \mathfrak{F}_t\}$ *is a martingale difference array, there exists* $\epsilon > 0$ *such that* $T^{-1}\sum_{t=1}^T E[||s_t(\theta_0)||^{2+\epsilon}|\mathfrak{F}_{t-1}] = O_p(1)$ *and* $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]} E[s_t(\theta_0)s_t(\theta_0)'|\mathfrak{F}_{t-1}] - \int_0^\lambda \Gamma(l)dl|| \xrightarrow{p} 0$*.*

Condition 1 is a set of fairly standard high level assumptions on the 'forecast error decomposition'-part of the likelihood. (DIFF) assumes existence of two derivatives. (ID) is similar to the global identification condition assumed in Schervish (1995), page 436, somewhat strengthened to ensure that even a slightly perturbed evaluation of the likelihood at parameter values different from θ_0 still yields a 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 this assumption. Part (iii) allows the information accrual to vary over the sample, and $\Gamma(\lambda)$ describes the average information at time $t = [\lambda T]$. If $h_t(\theta_0)$, $t = 1, \cdots, T$ is positive semidefinite almost surely, part (ii) of (LLLN) is implied by part (iii). (MDA) assumes the sequence of scores to constitute a martingale difference array with slightly more than two conditional moments, with an average conditional variance of $\Gamma(\lambda)$ at time $t = [\lambda T]$. Whenever the relevant conditional moments exist, $\{s_t(\theta_0\}, \mathfrak{F}_t\}$ and $\{s_t(\theta_0)s_t(\theta_0)' - h_t(\theta_0), \mathfrak{F}_t\}$ are martingale difference arrays by construction—see Hall and Heyde (1980), Chapter 6.2. Phillips and Ploberger (1996) and Li and Müller (2006) make very similar assumptions to (LLLN) and (MDA).

Now consider an unstable version of this parametric model, with time varying parameter $\theta_t = \theta + \delta_t$, $t = 1, \cdots, T$, so that the density of the data y_T becomes

$$f_T(\theta, \delta) = \prod_{t=1}^T f_{T,t}(\theta + \delta_t), \quad \theta + \delta_t \in \Theta \text{ for } t = 1, \cdots, T \tag{7}$$

where θ and δ_t are $k \times 1$ and $\delta = (\delta_1', \cdots, \delta_T')' \in \mathbb{R}^{Tk}$. Alternative estimators of $\{\theta + \delta_t\}_{t=1}^T$, or generally actions, 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 topological space 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_T 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

$$WAR(\hat{a}) = \int_\Theta w(\theta) \int_{\mathcal{V}_T(\theta)} r(\theta, \delta, \hat{a}) dQ_T(\delta) d\theta \tag{8}$$

The weighting functions w and Q_T describe the importance attached to alternative true parameter paths in the overall risk calculations: The weight function w attaches different weights to the benchmark value θ, and Q_T describes the focus on deviations from this baseline value. In the parametrization $\{\theta_t\}_{t=1}^T = \{\theta+\delta_t\}_{t=1}^T$, the average $T^{-1}\sum \delta_t$ and θ are obviously not uniquely identified. The same weighted average risk criterion may thus be expressed by different choices of w and Q_T. The parametrization is useful because the weighting schemes analyzed in this paper assume different asymptotic properties of Q_T and w as follows.

Condition 2 *(GS) The weight function Q_T is the distribution of $\{T^{-1/2}G(t/T)\}_{t=1}^T$, where G is a $k \times 1$ zero mean Gaussian semimartingale on the unit interval with covariance kernel $E[G(r)G(s)'] = \kappa_G(r, s)$. There exists a finite set of numbers $\tau = \{0, \tau_1, \cdots, \tau_q\} \subset [0, 1]$ such that $||\partial^2\kappa_G(r, s)/\partial r \partial s||$ and $||\partial^2\kappa_G(r, s)/\partial r^2||$ are bounded when $r, s \notin \tau$ and $r \neq s$, κ_G admits bounded left and right derivatives with respect to r for all $r = s \in [0, 1] \backslash \tau$, and $\partial\kappa_G(r, s)/\partial r$ is bounded for $r \in [0, s)\backslash\tau$ and $s \in \tau$.*

(CNT) The weight function w does not depend on T and w is continuous at θ_0.

Under Condition 2 (GS), the weight function Q_T focusses on persistent paths of relatively small variability.

Gaussian processes that satisfy the differentiability assumptions on their kernel are almost surely continuous for all $s \in [0, 1]\backslash\tau$ by Kolmogorov's continuity theorem, with τ_i, $i = 1, \cdots, q$, describing fixed break dates. This concentration on persistent parameter paths drives the derivation of the asymptotic equivalence results below, and it is appealing in many applications, as parameter instability is typical thought of as a low frequency phenomenon.

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. As discussed in Section 2 above, a popular choice in applied work has been the assumption that parameters vary as a Gaussian Random Walk, which may be achieved by setting G equal to $G(\cdot) = \Upsilon^{1/2}W(\cdot)$, where W is a $k \times 1$ standard Wiener process. Random walk parameter variability that only occurs in, say, the first half of the sample is achieved by letting $G(s) = \mathbf{1}[s \leq 1/2]\Upsilon^{1/2}W(s) + \mathbf{1}[s > 1/2]\Upsilon^{1/2}W(1/2)$. An assumption of slowly mean reverting parameters can be expressed by letting G be a stationary Ornstein-Uhlenbeck process, etc.

Under Condition 2 (GS), the weighted average risk criterion (8) focusses on parameter paths whose variability is of order of magnitude $T^{-1/2}$. This choice 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 of the magnitude of $\{\delta_t\}$ under Q_T 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 below based on the sequence of weights as described Condition 2 (GS) becomes relevant to the small sample problem where there is in fact little information about the parameter evolution.

The order of magnitude $T^{-1/2}$ for δ_t under Condition 2 (GS) corresponds to the local neighborhood in which efficient stability tests have nontrivial asymptotic power. The null hypothesis of a stability test is that the parameter path $\{\theta_t\}_{t=1}^T = \{\theta + \delta_t\}_{t=1}^T$ is constant, i.e.

$$H_0 : \delta_t = 0 \quad \text{for } t = 1, \cdots, T \tag{9}$$

against the alternative that the parameter is time varying. For the development of optimal parameter stability tests, it makes sense to restrict the parameter paths under the alternative such that the difference to the corresponding stable model is the time variability of the path, rather than a different average value of the path. The appropriate restriction is achieved by the multivariate Gaussian measure Q_T^* of $\{T^{-1/2}(G(t/T) - (\sum_{s=1}^T \Gamma(s/T))^{-1} \sum_{s=1}^T \Gamma(s/T)G(s/T))\}_{t=1}^T$. When information accrual is constant, that is $\Gamma(s) = H$ for all $s \in [0,1]$, then the restriction amounts to a demeaning of δ_t, such that $\sum \delta_t = 0$ a.s. under Q_T^*. In the general case, the restriction forces $\sum \Gamma(t/T)\delta_t = 0$, so that the information weighed parameter path deviations sum to zero, just as in the efficient tests derived by Andrews and Ploberger (1994). Intuitively, a model with time varying parameter is closest to the stable model with a parameter that is the information weighted average of the parameter path.

Possibly randomized parameter stability tests φ_T are measurable functions from the data to the interval $[0, 1]$, where $\varphi_T(y_T)$ indicates the probability of rejecting the null hypothesis of parameter stability when observing y_T. Tests of the same size can then usefully be compared by considering their Weighted Average Power

$$WAP(\varphi_T) = \int_{\mathcal{V}_T(\theta)} \int f_T(\theta_0, \delta)\varphi_T d\mu_T dQ_T^*(\delta) \tag{10}$$

as suggested by Andrews and Ploberger (1994). While θ_0 is typically unknown, we show below that there exists a feasible test φ_T^* that asymptotically maximizes this weighted average power.

With the weighting of parameter paths specified as the distribution of a Gaussian process, the problem of finding weighted average risk minimizing actions essentially becomes a nonlinear smoothing exercise. The weighted average risk minimizing decision is to choose the action a that minimizes

$$\frac{\int_\Theta w(\theta) \int_{\mathcal{V}_T(\theta)} f_T(\theta, \delta) L_T(\theta, \delta, a) dQ_T(\delta) d\theta}{\int_\Theta w(\theta) \int_{\mathcal{V}_T(\theta)} f_T(\theta, \delta) dQ_T(\delta) d\theta} \tag{11}$$

for each data y_T. With the weighting functions normalized to integrate to unity, this is simply Bayes Rule for minimizing Bayes risk (11), 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 2.

A large literature has developed around numerically finding exact posterior distributions in nonlinear filtering/smoothing problems, usually by Monte Carlo simulation techniques. 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.

Note that Condition 1 makes assumptions about the stable model only, that is on its behavior when the parameter path is constant. Clearly, with a focus on the problem of estimating the parameter path, we need to argue for the accuracy of approximations also when the true data generating process has time varying parameters. In general, most models with time varying parameters generate nonstationary data, to which standard asymptotic results are not easily applicable. In a Vector Autoregressive Regression model, for instance, parameter instabilities lead to highly complicated interactions between the evolution of the lagged variables and the unstable parameters. Our approach is thus to derive asymptotic results for unstable models as an implication of the *contiguity* of models with time varying parameters of order $T^{-1/2}$ to the corresponding stable model, similar to Andrews and Ploberger (1994), Phillips and Ploberger (1996), Elliott and Müller (2006) and Li and Müller (2006). The following Lemma follows from Lemma 1 of Li and Müller (2006) and the additional discussion in their appendix.

Lemma 1 *Let $\pi_0 : [0,1] \mapsto \mathbb{R}^k$ be a piece-wise continuous function with at most a finite number of discontinuities. Under Condition 1 the sequence of densities $\prod_{t=1}^T f_{T,t}(\theta_0, T^{-1/2}\pi_0(t/T))$ is contiguous to the sequence $f_T(\theta_0, 0)$. Furthermore, the two sequences of densities $\int_{\mathcal{V}_T(\theta_0)} f_T(\theta_0,\delta)dQ_T(\delta)/\int_{\mathcal{V}_T(\theta_0)} dQ_T(\delta)$ and $\int_{\mathcal{V}_T(\theta_0)} f_T(\theta_0,\delta)dQ_T^*(\delta)/\int_{\mathcal{V}_T(\theta_0)} dQ_T^*(\delta)$ are contiguous to the sequence $f_T(\theta_0, 0)$.*

The main result of the paper is the following Theorem.

Theorem 1 *Let the sequence of positive definite matrices $\{\tilde{h}_t\}_{t=1}^T = \{\tilde{h}_{T,t}\}_{t=1}^T$ satisfy*

$$\sup_{\lambda\in[0,1]} \left\| T^{-1}\sum_{t=1}^{[\lambda T]} \tilde{h}_t - \int_0^\lambda \Gamma(s)ds \right\| \xrightarrow{p} 0 \tag{12}$$

in the stable model with parameter θ_0.

(i) Assume that the decision $\hat{a}^*$ minimizes weighted average risk with weights as in Condition 2 or a flat weighting of θ and the weight function Q_T on δ in the pseudo model

$$s_t(\hat{\theta}) + \tilde{h}_t\hat{\theta} = \tilde{h}_t(\delta_t + \theta) + \nu_t, \quad \nu_t \sim \textit{independent}\ \mathcal{N}(0,\tilde{h}_t), \quad t = 1,\cdots,T. \tag{13}$$

If Condition 1 and (12) hold for almost all θ_0 in the support of w, then for all $\hat{a}$, $\liminf_{T\to\infty}[WAR(\hat{a}) - WAR(\hat{a}^*)] \geq 0$.

(ii) Let $\tilde{Q}_T^*$ be the distribution of $\{T^{-1/2}G(t/T) - T^{-1/2}(\sum_{s=1}^T \tilde{h}_s)^{-1}\sum_{s=1}^T \tilde{h}_s G(s/T)\}_{t=1}^T$ (induced by G), and let φ_T^* be the level α test of (9) that maximizes weighted average power with respect to the weighting function $\tilde{Q}_T^*$ in the pseudo model

$$s_t(\hat{\theta}) = \tilde{h}_t\delta_t + \nu_t, \quad \nu_t \sim \textit{independent}\ \mathcal{N}(0,\tilde{h}_t), \quad t = 1,\cdots,T. \tag{14}$$

Then under Conditions 1 and 2, for any other test φ_T of (9) of asymptotic level α, $\liminf_{T\to\infty}[WAP(\varphi_T^*) - WAP(\varphi_T)] \geq 0$.

(iii) Under Condition 1, the total variation difference between the posterior distribution of (θ,δ) in model (7) with priors as in Condition 2 and the posterior distribution of (θ,δ) in the pseudo model (13) with either the same priors or with a flat prior on θ and prior Q_T on δ converges in probability to zero in both the stable model with parameter θ_0 and any unstable model that satisfies the condition of Lemma 1.

Theorem 1 asserts that asymptotically efficient decisions and tests are obtained from combining the sample information from pseudo models (13) and (14), respectively, with the weighting of Condition 2. Since both of these are Gaussian, the resulting distribution can be computed explicitly. Let e be the $Tk \times k$ matrix $e = (I_k, \cdots, I_k)'$, $D_{\tilde{h}} = \operatorname{diag}(\tilde{h}_1, \cdots, \tilde{h}_T)$, $\Sigma_\delta = E_\delta[\delta\delta']$, where E_δ denotes integration with respect to Q_T of Condition 2, $K = \Sigma_\delta(D_{\tilde{h}}\Sigma_\delta + I_{Tk})^{-1}$, $\hat{s} = (s_1(\hat{\theta})', \cdots, s_T(\hat{\theta})')'$ and

$$\Sigma = K + (I_{Tk} - KD_{\tilde{h}})e(e'D_{\tilde{h}}e - e'D_{\tilde{h}}KD_{\tilde{h}}e)^{-1}e'(I_{Tk} - D_{\tilde{h}}K). \tag{15}$$

Note that with $\delta \sim \mathcal{N}(0, \Sigma_\delta)$ and the measurements $Y_t = \tilde{h}_t\delta_t + \nu_t$, $\nu_t \sim$ independent $\mathcal{N}(0, \tilde{h}_t)$, $t = 1, \cdots, T$, the distribution of δ conditional on the measurements $Y = (Y_1', \cdots, Y_T')$ and $D_{\tilde{h}}$ is $\delta|(Y, D_{\tilde{h}}) \sim \mathcal{N}(KY, K)$. The second term in the definition of Σ results from the uncertainty concerning the baseline value θ. The matrix Σ remains the same if Σ_δ is substituted by the covariance matrix of δ under $\tilde{Q}_T^*$, as defined in Theorem 1 (ii).[1]

Theorem 2 *Let Π be the distribution $\mathcal{N}(e\hat{\theta} + \Sigma\hat{s}, \Sigma)$.*

(i) The decision $\hat{a}^*$ that minimizes expected risk relative to the distribution $e\theta + \delta \sim \Pi$ for each y_T minimizes weighted average risk in the pseudo model (13) with a flat weighting on θ.

(ii) A test that rejects for large values of $\hat{s}'\Sigma\hat{s}$ is the optimal stability test in the pseudo model (14), and under Conditions 1 and 2

$$\hat{s}'\Sigma\hat{s} \Rightarrow 2\ln\left(\frac{E_G \exp[\int G^*(s)'\Gamma(s)^{1/2}dW^*(s) - \frac{1}{2}\int G^*(s)'\Gamma(s)G^*(s)ds]}{E_G \exp[-\frac{1}{2}\int G^*(s)'\Gamma(s)G^*(s)ds]}\right)$$

under the null hypothesis, where $G^*(s) = G(s) - (\int \Gamma(\lambda)d\lambda)^{-1}\int \Gamma(\lambda)G(\lambda)d\lambda$, the standard $k \times 1$ Wiener process W^* is independent of G and E_G denotes integration with respect to the probability measure of G.

(iii) The posterior distribution of $e\theta + \delta$ under a flat prior on θ in the pseudo model (13) is given by Π.

Comments:

1. Part (i) of Theorem 1 establishes that for arbitrary bounded loss functions, the decision that minimizes weighted average risk in the Gaussian pseudo model (13) is also asymptotically optimal in the true model. As shown in part (i) of Theorem 2, this amounts to finding the risk minimizing action relative to a multivariate Gaussian distribution for the parameter path.

[1] This follows from Theorem 2 (i): combined with the flat weighting on θ, all weighting functions for δ_t that imply the same weighting for $\{\delta_t - T^{-1}\sum_{s=1}^T \delta_s\}_{t=1}^T$ yield the same overall weighting function for $\{\theta + \delta_t\}_{t=1}^T$.

Note that loss may be defined arbitrarily (subject to the bounding condition) for parameter values outside Θ, allowing the problem in the pseudo model to be made entirely spherical. For the wide range of bounded bowl-shaped loss functions for which one would choose the posterior mean in a Gaussian model, an asymptotically efficient parameter path estimator is hence given by $e\hat{\theta} + \Sigma\hat{s}$. Note that such loss functions include those that consider a weighted average of symmetric losses incurred by estimation errors in the parameter value, such as

$$L_T(\theta, \delta, a) = \sum_{t=1}^{T} q_{T,t} L_0(T(\theta + \delta_t - a_t)' W_L(\theta + \delta_t - a_t)) \tag{16}$$

where $a = (a_1', \cdots, a_T')' \in \mathbb{R}^{Tk}$, $\inf_{t \leq T} q_{T,t} \geq 0$, $\sum_{t=1}^{T} q_{T,t} = 1$, W_L is a nonnegative definite $k \times k$ matrix and $L_0 : [0, \infty) \mapsto [0, \bar{L}]$ is a monotonically nondecreasing, bounded function with $L_0(0) = 0$. The scaling by T in (16) ensures that the loss does not become trivial as $T \to \infty$ even for good path estimators, although Theorems 1 and 2 remain true without this scaling. This class of loss functions (16) contains the special case where one only cares about the parameter at time T, i.e. $q_{T,T} = 1$ and $q_{T,t} = 0$ for all $t < T$, which arises 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 Gaussian pseudo model. This typically represents a dramatic computational simplification.

2. Part (ii) of Theorems 1 and 2 spell out the implications of the approximation for efficient tests of the null hypothesis of parameter stability (9). Part (i) of Theorem 2 shows that under symmetric loss, the asymptotically efficient parameter path estimator is $e\hat{\theta} + \Sigma\hat{s}$ with an asymptotic uncertainty described by a zero mean multivariate normal with covariance matrix Σ. The asymptotically efficient test statistic $\hat{s}'\Sigma\hat{s} = (\Sigma\hat{s})'\Sigma^{+}(\Sigma\hat{s})$, where Σ^{+} denotes a general inverse, is recognized to be of the usual Wald form: Efficient instability tests are based on a quadratic form in the efficient estimator of the instability. Efficient estimation and testing in (potentially) unstable models are hence unified in one coherent framework. This ensures coherence between the stability test and the path estimator, as $\hat{s}'\Sigma\hat{s}$ can only be large if the path estimator $e\hat{\theta} + \Sigma\hat{s}$ shows substantial variation.

3. Part (iii) of Theorems 1 and 2 describe the approximation result in Bayesian terms: The posterior distribution of the parameter path $e\theta + \delta$ comes arbitrarily close to the Tk dimensional multivariate normal distribution $\mathcal{N}(e\hat{\theta} + \Sigma\hat{s}, \Sigma)$. This is a considerably 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. With $G(s) = 0$, so that $\Sigma_\delta = K = 0$, Σ becomes

$\Sigma = e(e'D_{\tilde{h}}e)^{-1}e'$, and one recovers the standard result that the posterior distribution of θ converges to $\mathcal{N}(\hat{\theta}, T^{-1}\tilde{H}^{-1})$ where $\tilde{H} = T^{-1}\sum \tilde{h}_t \xrightarrow{p} \int \Gamma(\lambda)d\lambda$, the average information.

In practice, part (iii) of Theorem 1 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 an iterative algorithm for generating random variables with distribution $\mathcal{N}(\Sigma\hat{s}, \Sigma)$ for the special case where G is a $k \times 1$ Wiener process.

4. The asymptotic distribution of the asymptotically efficient test statistic $\hat{s}'\Sigma\hat{s}$ is provided in Theorem 2 (ii). This distribution is nonstandard and depends on the weighting function G and the evolution of the information Γ. Even with Γ known, a simulation based on this expression is quite cumbersome due to the integration over the measure of G. The usefulness of Theorem 2 (ii) is that it shows the existence of an asymptotic distribution. It thus suffices to consider a computationally convenient stable model that has the same asymptotic distribution, such as the stable Gaussian location model $y_t = \tilde{h}_t\theta + Z_t$, $t = 1, \cdots, T$ with Z_t independent and distributed $\mathcal{N}(0, \tilde{h}_t)$. The limiting distribution of $\hat{Z}'\Sigma\hat{Z}$ with $\hat{Z} = (\hat{Z}_1', \cdots, \hat{Z}_T')'$ and $\hat{Z}_t = Z_t - \tilde{h}_t(\sum_{s=1}^T \tilde{h}_s)^{-1}\sum_{s=1}^T \tilde{h}_s Z_s$ is therefore the same as the asymptotic null distribution of $\hat{s}'\Sigma\hat{s}$, for data drawn both from the stable model and under all local alternatives for which Lemma 1 implies (12) to also hold.[2] Asymptotically justified critical values of the test statistic $\hat{s}'\Sigma\hat{s}$ might hence be obtained by considering the empirical distribution of sufficiently many draws from the distribution of $\hat{Z}'\Sigma\hat{Z}$, similar to the approach of Hansen (1996). In the appendix, we provide an iterative algorithm for computing $\hat{s}'\Sigma\hat{s}$ (and $\hat{Z}'\Sigma\hat{Z}$) that does not require inversion of $Tk \times Tk$ matrices when G is a Wiener process.

5. In contrast to Theorem 1 (ii), part (i) requires Condition 1 to hold for almost all θ_0 in the support of w. This restriction can be relaxed for general decision problems that only involve δ, the deviations of the parameter path from its baseline value, such as assessing their shape or size. If this is formalized with the same weighting function for δ as employed in the testing problem (10), i.e.

$$\widetilde{WAR}(\hat{a}) = \int_{\mathcal{V}_T(\theta)} \int f_T(\theta_0, \delta)\tilde{L}_T(\delta, \hat{a})d\mu_T dQ_T^*(\delta)$$

where $\tilde{L}_T : \mathbb{R}^{Tk} \times \mathbb{A}_T \mapsto [0, \bar{L}]$, one obtains asymptotically efficient feasible decisions based

[2]Formally, this follows from replacing $\hat{s}$ and s_0 by $\hat{Z}$ and $Z = (Z_1', \cdots, Z_T')'$, respectively, in the derivation of the asymptotic null distribution in Theorem 2 (ii).

on the pseudo model (14).[3] Asymptotically efficient decisions thus minimize expected risk relative to the distribution $\delta \sim \mathcal{N}(\Sigma\hat{s}, \Sigma)$. In particular, under a bowl-shaped symmetric loss function, the asymptotically efficient estimator of δ is $\Sigma\hat{s}$. This approximation only requires Condition 1 to hold for the θ_0 that generates the data.

6. The approximation results in Theorems 1 and 2 hold for any choice of positive definite sequences $\{\tilde{h}_t\}_{t=1}^T$ that satisfy (12) in the stable model. For models with almost surely positive definite $h_t(\theta)$, $t = 1, \cdots, T$, $\theta \in \Theta$, a natural choice is given by $\tilde{h}_t = h_t(\hat{\theta})$, which satisfies (12) under Condition 1, as shown in Lemma 2 (vi) in the appendix. One might gain some small sample approximation accuracy by iterating with $\tilde{h}_t = h_t(\hat{\theta}_t^1)$, $t = 1, \cdots, T$, where $\hat{\theta}_t^1$ is a preliminary path estimator, although for large enough T, all choices for $\tilde{h}_t$ satisfying (12) yield equivalent results.

From a computational point of view, a particular convenient choice would be to rely on the outer product of scores, $\tilde{h}_t = s_t(\hat{\theta})s_t(\hat{\theta})'$, which satisfies (12) under Condition 1 (see Lemma 2 (v) in the appendix), so that no second order derivatives of the log-likelihood are required. With this choice, $\tilde{h}_t$ is of course singular when $k > 1$. One would be formally justified in invoking Theorems 1 and 2 with $\tilde{h}_t = s_t(\hat{\theta})s_t(\hat{\theta})' + \varkappa_T I_k$, where $\varkappa_T$ is any sequence of positive real numbers converging to zero, at an arbitrarily fast rate. Note, however, that Σ is a continuous function of the eigenvalues of $\{\tilde{h}_t\}_{t=1}^T$, with a well defined limit as $\varkappa_T \to 0$ for fixed T. One might thus drop the additional correction $\varkappa_T I_k$ and set $\tilde{h}_t = s_t(\hat{\theta})s_t(\hat{\theta})'$ in the definition of Σ in (15) without affecting the validity of Theorem 2. (Note, however, that Theorem 1 is false for singular $\tilde{h}_t$; in general, the pseudo model (13) with, say, $\tilde{h}_t = s_t(\hat{\theta})s_t(\hat{\theta})'$ leads to a different posterior distribution than the limit of the posterior distributions for $\tilde{h}_t = s_t(\hat{\theta})s_t(\hat{\theta})' + \varkappa_T I_k$ as $\varkappa_T \to 0$.)

7. For certain applications it makes sense to make the scale of the weighting function in the estimation (8) and testing problems (10) a function of the information Γ. In a testing context, for instance, it often attractive to choose G such that alternatives that are equally difficult to detect receive a similar weight, as in Wald (1943) and, conditional on the break date, in Andrews and Ploberger (1994). Typically, of course, Γ is unknown, and needs to be estimated from the data. Optimal decisions and tests from the pseudo models (13) and (14) with respect to an estimated weighting function generally continue to be asymptotically optimal decisions in terms of (8) and (10), i.e. with respect to the data independent weighting functions described in Condition 2.

[3]This follows directly from combining the arguments in the proof of Theorem 1 (i) with the results employed in the proof of Theorem 1 (ii).

Theorem 3 *Suppose* $\{\hat{\Lambda}_{T,t}\}_{t=1}^T$ *are nonsingular* $k \times k$ *statistics such that* $\sup_{t\leq T} ||\hat{\Lambda}_{T,t} - I_k|| \xrightarrow{p} 0$ *and* $\sum_{t=2}^T ||\hat{\Lambda}_{T,t} - \hat{\Lambda}_{T,t-1}|| \xrightarrow{p} 0$ *in the stable model with parameter* θ_0. *Then part (ii) of Theorem 1 also holds for* $\tilde{Q}_T^*$ *replaced by the distribution of* $\{T^{-1/2}\hat{\Lambda}_{T,t}G(t/T) - T^{1/2}(\sum_{s=1}^T \tilde{h}_s)^{-1}\sum_{s=1}^T \tilde{h}_s\hat{\Lambda}_{T,s}G(s/T)\}_{t=1}^T$ *(induced by* G*). Furthermore, if* $\sup_{\theta\in\Theta,\delta\in\mathbb{R}^{Tk},a\in\mathbb{A}_T} |L_T(\theta, \mathrm{diag}(\Lambda_{T,1}, \cdots, \Lambda_{T,T})\delta, a) - L_T(\theta, \delta, a)| \to 0$ *for all sequences* $\{\Lambda_{T,t}\}_{t=1}^T$ *satisfying* $\sup_{t\leq T} ||\Lambda_{T,t} - I_k|| \xrightarrow{p} 0$ *and* $\sum_{t=2}^T ||\Lambda_{T,t} - \Lambda_{T,t-1}|| \to 0$ *as* $T \to \infty$, *then also part (i) of Theorem 1 holds for* Q_T *replaced by the distribution of* $\{T^{-1/2}\hat{\Lambda}_{T,t}G(t/T)\}_{t=1}^T$ *(induced by* G*).*

In a typical application of Theorem 3, suppose one aims at computing the asymptotically efficient test for a Condition 2 weighting function with $G(\cdot) = c\bar{\Gamma}^{-1/2}W(\cdot)$, where c is a known scalar constant, but the average information $\bar{\Gamma} = \int_0^1 \Gamma(\lambda)d\lambda$ is not known. Then Theorem 3 shows that this test may be computed from the pseudo model (14) with an estimated weighting function that corresponds to the distribution of $c\hat{\bar{\Gamma}}^{-1/2}W(\cdot) = c\hat{\bar{\Gamma}}^{-1/2}\bar{\Gamma}^{1/2}G(\cdot)$, i.e. based on the statistic $\hat{s}'\Sigma\hat{s}$ where Σ_δ in the definition (15) of Σ has i, jth $k \times k$ block equal to $T^{-2}c^2\sum_{t=1}^{\min(i,j)} \hat{\bar{\Gamma}}^{-1}$, as long as $\hat{\bar{\Gamma}} \xrightarrow{p} \bar{\Gamma}$ under θ_0 stable. In the more general case where $G(\cdot) = \Omega(\cdot)G_0(\cdot)$ with G_0 a known Gaussian process and $\Omega : [0, 1] \mapsto \mathbb{R}^{k\times k}$ an unknown fixed and nonsingular matrix function, Theorem 3 requires beyond consistency that the scaled estimation error $\hat{\Lambda}_{T,t} = \hat{\Omega}_{T,t}\Omega(t/T)^{-1}$ is smooth by imposing $\sum_{t=2}^T ||\hat{\Lambda}_{T,t} - \hat{\Lambda}_{T,t-1}|| \xrightarrow{p} 0$. This condition is typically satisfied for parametric estimators of Ω when Ω is of bounded variation, such as, for example, when Ω is a linear trend of unknown slope or when Ω is a step function with known step locations.

Moreover, optimal decisions from the pseudo model typically retain their weighted average risk (8) optimality under such estimated weights, such as the path estimator $e\hat{\theta} + \Sigma\hat{s}$ under the class of loss functions (16) when L_0 is Lipschitz continuous. The restriction of the loss functions in the second claim of Theorem 3 is necessary to rule out a somewhat pathological focus of L_T on the scale of the weighting function for δ.[4]

8. Much applied work is based on the special case where the prior or weighting function of a time varying parameter is a Gaussian random walk, such that $G(\cdot) = \Upsilon^{1/2}W(\cdot)$ for some positive semidefinite matrix Υ and standard Wiener process W; see the citations in Section 2. The Markovian structure of the Wiener process enables the application of an iterative Kalman smoothing algorithm for the computation of the path estimator $e\hat{\theta} + \Sigma\hat{s}$ and the test statistic

[4]For example, with $G(s) = W(s)$ and $\Lambda_{T,t} = (1+T^{-1/4})I_k$, $L_T(\theta, \delta, a) = (T^{1/2}\,\mathrm{tr}(T\sum(\Delta\delta_t)(\Delta\delta_t)' - I_k))^2 \wedge 1$, $\lim_{T\to\infty}\int L_T(\theta, \delta, a)dQ_T(\delta) \neq \lim_{T\to\infty}\int L_T(\theta, (1 + T^{-1/4})\delta, a)dQ_T(\delta)$.

$\hat{s}'\Sigma\hat{s}$ that avoids matrix computations of dimension $Tk \times Tk$. We provide such an algorithm in the appendix, which also takes care of the impact of the flat weighting of θ in the smoothing, along similar lines as Rosenberg (1973) and Jong (1991).

A number of previous papers have considered parameter stability tests against random walk-type alternatives: Nyblom (1989) derives locally best tests against general martingale variability in the parameters for general likelihood models, Shively (1988a, 1988b) considers small sample tests in a linear regression model, and Elliott and Müller (2006) derive asymptotic results for point optimal parameter instability tests in linear regression models for a class of weighting functions that includes the Gaussian random walk case. The contribution of Theorems 1 and 2 with respect to this literature is the generalization of the point optimal tests to general likelihood models, including nonstationary models with, say, a time trend. The degree of generality of the results here concerning parameter stability tests is similar to those of Andrews and Ploberger (1994), but for a different type of weighting functions.

Elliott and Müller (2006) show that efficient tests for a Gaussian random walk in the parameters and efficient tests for a single break at unknown date have asymptotic power that is roughly comparable no matter what the true alternative is; the efficient tests for the Gaussian random walk have the advantage that they avoid the need for trimming the break dates away from the beginning and end of the sample, and their computational convenience, at least compared to efficient tests for more than one potential break.

9. An important special case arises when the information accrual is constant, i.e. $\Gamma(s) = H$ for all $s \in [0,1]$ in Condition 1. This holds in particular for all stationary models that satisfy Condition 1. In that case, one might choose $\tilde{h}_t = \hat{H}$, $t = 1, \cdots, T$ with $\hat{H} \xrightarrow{p} H$ in the stable model in an application of Theorem 1, and the pseudo model (13) becomes (3).

When $\Gamma(s) = H$ for all $s \in [0,1]$ and $G(\cdot) = \Upsilon^{1/2}W(\cdot)$, a theoretically appealing choice for Υ is $\Upsilon = c^2 H^{-1}$ for some scalar c. This choice equates the degree of uncertainty about the time variation of δ_t in any given direction (in $\mathbb{R}^k$) with the average sample information about that direction, as under Condition 1, 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 for Υ that yields asymptotic results that do not depend on a particular parametrization. Nyblom (1989), Stock and Watson (1998) and Elliott and Müller (2006) argue for the same choice for their testing procedures. By an application of Theorem 3, a consistent estimator of $\hat{H}$ is sufficient to implement asymptotically efficient tests and most weighted average risk minimizing estimators under the weighting $G(\cdot) = cH^{-1/2}W(\cdot)$ for a given c.

When only the first $p \leq k$ elements β of θ are (potentially) time varying, $\Upsilon =$

$\operatorname{diag}(I_p, 0_{k-p})c^2 H^{-1} \operatorname{diag}(I_p, 0_{k-p})$ remains an attractive choice, as it yields asymptotic results that remain invariant to reparametrizations of β. The algorithm described in Section 2 of this paper exploits the additional computational simplifications when $\tilde{h}_t = \hat{H}$, $t = 1, \cdots, T$. In particular, the smoothing algorithm and formulas provided in Section 2 follow by combining our results with those of Elliott and Müller (2006): applying the matrix identity (21) in the appendix, Σ in (15) becomes $(I_T - G_c) \otimes (\operatorname{diag}(I_p, 0_{k-p})c^2 \hat{H}^{-1} \operatorname{diag}(I_p, 0_{k-p}))$ in their notation, such that the asymptotic distribution of $-\hat{s}'\Sigma\hat{s}$ simplifies to the one given in their Theorem 4, and the expression for $\kappa_t(c)$ in Section 2 follows from their proof of Lemma 6, as the t, tth element of $I_T - G_c$ equals $\iota_t'(I_T - G_c)\iota_t$, where ι_t is the $T \times 1$ vector with a one as the tth element and zeros elsewhere.

10. For some purposes, it makes sense to consider weighting functions that are more agnostic about the magnitude and/or form of the parameter instability than is possible under Condition 2. One way to achieve this without foregoing the computational advantages of a Gaussian weighting function is to consider weighting functions (or priors) for δ that are a weighted average of distributions of different Gaussian processes. The following Theorem shows how parts (i) and (iii) of Theorems 1 and 2 need to be adapted in the case of such a finite mixture.

Theorem 4 *Let G_i, $i = 1, \cdots, n_G$ be processes satisfying Condition 2 (GS). If Q_T is the distribution of the mixture of $\{T^{-1/2}G_i(t/T)\}$ with mixing probabilities p_i, then parts (i) and (iii) of Theorems 1 and 2 hold with Π replaced by the mixture of n_G multivariate normal distributions $\mathcal{N}(e\hat{\theta} + \Sigma_i\hat{s}, \Sigma_i)$ with mixing probabilities proportional to*

$$\tilde{w}_i = p_i|D_{\tilde{h}}\Sigma_{\delta(i)} + I_{Tk}|^{-1/2}|e'D_{\tilde{h}}e - e'D_{\tilde{h}}K_iD_{\tilde{h}}e|^{-1/2}\exp[\tfrac{1}{2}\hat{s}'\Sigma_i\hat{s}], \quad i = 1, \cdots, n_G, \qquad (17)$$

where K_i, $\Sigma_{\delta(i)}$ and Σ_i are defined as K, Σ_δ and Σ in (15) with Σ_δ replaced by $\Sigma_{\delta(i)}$, the covariance matrix of $T^{-1/2}(G_i(1/T)', G_i(2/T)', \cdots, G_i(1)')'$ for $i = 1, \cdots, n_G$.

Theorem 4 is a simple consequence of the fact that the Gaussian pseudo model (13) remains an accurate approximations of the sample information for each of the n_G weighting functions, such that the likelihood ratios can be explicitly computed. The weighted average risk minimizing parameter path estimator under mixture weightings generally depends much more on the loss function than in the single Gaussian process case, as mixture of normal distributions are not generally symmetric around their mean. Under truncated quadratic loss (16) with $L_0(x) = \min(x, \bar{L})$, the weighted average risk minimizing path estimator converges to $\sum_{i=1}^{n_G} \tilde{w}_i\Sigma_i\hat{s} / \sum_{i=1}^{n_G} \tilde{w}_i$ as $\bar{L} \to \infty$. In the appendix, we provide a computational convenient way of computing the determinants appearing in (17) when the Gaussian processes G_i of Theorem 4 are Wiener processes of some covariance matrix. If in addition information accrual is linear, i.e. $\Gamma(\cdot) = H$ in Condition 1, the determinants are computed in closed from in Elliott and Müller (2006), and these expressions are given in Section 2 above.

11. Theorems 1–4 make asymptotic efficiency claims about estimators and tests in correctly specified parametric models. It is plausible that the resulting test statistics remain asymptotically valid for a larger class of data generating processes, as long as the score remains a valid moment condition, and the variance estimator is of the outer product form $\tilde{h}_t = s_t(\hat{\theta})s_t(\hat{\theta})'$, $t = 1, \cdots, T$. Elliott and Müller (2006) provide such results in linear time series regressions for the special case where $\Gamma(\cdot) = H$ and $G(\cdot) = \Upsilon^{1/2}W(\cdot)$. It seems likely that similar statements hold true for the wider class of efficient tests considered here, but we leave such extensions to future research.

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 substantial numerical and computational complications.

This paper considers a general likelihood model and focusses on parameter instabilities of a magnitude that are nontrivial to detect, which seems a relevant part of the parameter space for many instabilities economists care about. The main contribution is an asymptotically justified approximation to the sample information about the time varying parameter, so that under a Gaussian weighting, weighted average risk minimizing path estimators and weighted average power maximizing parameter stability tests become straightforward to compute. We believe these results are not only of theoretical interest, but they add useful tool to the applied econometrician's toolbox: At least for a 'first look' at model with potentially unstable parameters, the procedures suggested here constitute an attractive alternative to numerical approximations to the exact solution, as 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

5.1 Iterative formulas for the path estimator and related statistics when $G(\cdot) = \Upsilon^{1/2}W(\cdot)$:

With $\hat{s}_t = s_t(\hat{\theta})$, compute

$$\begin{aligned}
\hat{a}_t &= \hat{a}_{t-1} + P_{t-1}(\tilde{h}_t P_{t-1} + I_k)^{-1}(\hat{s}_t - \tilde{h}_t \hat{a}_{t-1}) \\
\hat{A}_t &= \hat{A}_{t-1} + P_{t-1}(\tilde{h}_t P_{t-1} + I_k)^{-1}(\tilde{h}_t - \tilde{h}_t \hat{A}_{t-1}) \\
P_t &= P_{t-1} + T^{-2}\Upsilon - P_{t-1}(\tilde{h}_t P_{t-1} + I_k)^{-1}\tilde{h}_t P_{t-1}
\end{aligned}$$

for $t = 1, \cdots, T$ with $\hat{a}_0 = 0$, $\hat{A}_0 = 0$ and $P_0 = T^{-2}\Upsilon$. Further, compute

$$\begin{aligned}
\hat{b}_t &= \hat{a}_t + (I_k - T^{-2}\Upsilon P_t^{-1})(\hat{b}_{t+1} - \hat{a}_t) \\
\hat{B}_t &= \hat{A}_t + (I_k - T^{-2}\Upsilon P_t^{-1})(\hat{B}_{t+1} - \hat{A}_t) \\
R_t &= P_t - \Upsilon + (I_k - T^{-2}\Upsilon P_t^{-1})(R_{t+1} - P_t)(I_k - T^{-2}\Upsilon P_t^{-1})'
\end{aligned}$$

for $t = T-1, \cdots, 1$ with $\hat{b}_T = \hat{a}_T$, $\hat{B}_T = \hat{A}_T$ and $R_T = P_T - T^{-2}\Upsilon$. The tth $k \times 1$ block of $e\hat{\theta} + \Sigma\hat{s}$ is then given by

$$\hat{\theta} + \hat{b}_t - (I_k - \hat{B}_t)\left(\sum_{s=1}^{T} \tilde{h}_s (I_k - \hat{B}_s)\right)^{-1} \sum_{s=1}^{T} \tilde{h}_s \hat{b}_s$$

and the t,tth $k \times k$ block of Σ is given by $R_t + (I_k - \hat{B}_t)\left(\sum_s \tilde{h}_s (I_k - \hat{B}_s)\right)^{-1}(I_k - \hat{B}_t')$. Also, $\hat{s}'\Sigma\hat{s} = \sum_{t=1}^T \hat{s}_t'\hat{b}_t + (\sum_{t=1}^T \hat{s}_t'\hat{B}_t)\left(\sum_{t=1}^T \tilde{h}_t(I_k - \hat{B}_s)\right)^{-1}\sum_{t=1}^T \tilde{h}_t\hat{b}_t$, $|D_{\tilde{h}}\Sigma_\delta + I_{Tk}| = \prod_{t=1}^T |\tilde{h}_t P_{t-1} + I_k|$ and $|e'D_{\tilde{h}}e - e'D_{\tilde{h}}KD_{\tilde{h}}e| = |\sum_{t=1}^T \tilde{h}_t(I_k - \hat{B}_t)|$. To compute $\hat{Z}'\Sigma\hat{Z}$, replace $\hat{s}_t$ by $\hat{Z}_t$ throughout.

To generate a draw from $\mathcal{N}(e\hat{\theta} + \Sigma\hat{s}, \Sigma)$, one may proceed as follows: Draw $\tilde{b}_T \sim \mathcal{N}(\hat{a}_T, P_T - T^{-2}\Upsilon)$, and then draw iteratively for $t = T-1, \cdots, 1$

$$\tilde{b}_t \sim \mathcal{N}(\tilde{b}_{t+1} - T^{-2}\Upsilon P_t^{-1}(\tilde{b}_{t+1} - \hat{a}_t), T^{-2}\Upsilon - T^{-4}\Upsilon P_t^{-1}\Upsilon).$$

Draw $\tilde{d} \sim \mathcal{N}(0, \left(\sum_s \tilde{h}_s(I_k - \hat{B}_s)\right)^{-1})$ independent of $\{\tilde{b}_t\}_{t=1}^T$. Then $\{\hat{\theta} + \tilde{b}_t + (I_k - \hat{B}_t)\tilde{d}\}_{t=1}^T$ constitutes a draw from $\mathcal{N}(e\hat{\theta} + \Sigma\hat{s}, \Sigma)$.

If Υ is singular, then P_t^{-1} is to be replaced by the Moore-Penrose generalized inverse of P_t in the above computations.

5.2 Proofs

5.2.1 Notation

For notational ease, extend the domain of f_T by letting $f_T(\theta) = 0$ for $\theta \notin \Theta$, and let $s_t(\theta) = 0$ for $\theta \notin \Theta_0$, $t = 1, \cdots, T$.

The following notation is used in the following Lemmas and proofs:

- the $Tk \times k$ vector $e = (I_k, \cdots, I_k)'$
- the $k \times k$ matrices $\Gamma_t = \Gamma(t/T)$, $\tilde{H} = T^{-1} \sum \tilde{h}_t$ and $\hat{\Gamma} = T^{-1} \sum \Gamma_t$
- the $Tk \times Tk$ matrices $D_\Gamma = \text{diag}(\Gamma_1, \cdots, \Gamma_T)$, $D_{\tilde{h}} = \text{diag}(\tilde{h}_1, \cdots, \tilde{h}_T)$ and $F = T^{-1/2} F_0 \otimes I_k$, where F_0 is a $T \times T$ matrix with zeros above the main diagonal and ones elsewhere
- the $k \times 1$ vectors $u = T^{1/2}(\theta - \theta_0)$, $\hat{u} = T^{1/2}(\hat{\theta} - \theta_0)$, $\hat{s}_t = s_t(\hat{\theta})$, $t = 1, \cdots, T$ and $\bar{\delta} = \hat{\Gamma}^{-1} T^{-1} \sum_{t=1}^T \Gamma_t \delta_t$
- the $Tk \times 1$ vectors $\hat{s} = (\hat{s}_1', \cdots \hat{s}_T')'$ and $s_0 = (s_1(\theta_0)', \cdots, s_1(\theta_0)')'$
- the indicator functions $\mathcal{S}_T = \mathbf{1}[T^{1/2} \sup_{t \leq T} ||\delta_t|| < T^\eta]$, where η is defined in Condition 1 (ID) and we assume $\eta < 1/2$ without loss of generality and $\mathcal{A}_T = \mathbf{1}[||u|| < a_T]$ with $a_T \to \infty$ defined in Lemma 3 below
- the real valued functions $LR_T(u, \delta) = \frac{f_T(\theta_0 + T^{-1/2}u, \delta)}{f_T(\theta_0, 0)}$, $\widehat{LR}_T(u, \delta) = \exp[\sum \hat{s}_t' \delta_t - \frac{1}{2} \sum \delta_t' \tilde{h}_t \delta_t + T^{-1/2}(\hat{u} - u) \sum \tilde{h}_t \delta_t - \frac{1}{2} u' \tilde{H} u + \hat{u}' \tilde{H} u]$ and $\overline{LR}_T(\delta) = \exp[\sum \hat{s}_t' \delta_t - \frac{1}{2} \sum \delta_t' \tilde{h}_t \delta_t + \frac{1}{2}(T^{-1/2} \sum \delta_t' \tilde{h}_t) \tilde{H}^{-1} T^{-1/2} \sum \tilde{h}_t \delta_t]$
- the scalars $m_T = \int E_\delta w(\theta_0 + T^{-1/2}u) LR_T(u, \delta) du$, $\hat{m}_T = w(\theta_0) \int E_\delta \widehat{LR}_T(u, \delta) du$ and $M_T = E_\delta \prod_{t=1}^T \mathbf{1}[(\theta_0 + \delta_t) \in \Theta]$

5.2.2 Proofs of Theorems in the Main Text

The general strategy for the proof of Theorem 1 is as follows: Given Lemma 1, it suffices to prove convergences in probability for data generated under the stable model. All following probability calculations are thus made under the stable Condition 1 model, if not explicitly noted otherwise. We rely on a number of Lemmas that are stated and proven in Section 5.2.3 below.

We first establish part (iii) of Theorem 1, from which part (i) follows relatively easily. The main thrust of the proof of part (iii) is the argument that $\int E_\delta \left| w(\theta_0 + T^{-1/2}u) LR_T(u, \delta) - w(\theta_0) \widehat{LR}_T(u, \delta) \right| du$ converges in probability to zero. Lemma 3 (i) shows that replacing $LR_T(u, \delta)$ by $\mathcal{S}_T \mathcal{A}_T LR_T(u, \delta)$ in this expression induces a negligible approximation error. The approximation via Taylor series expansions is performed in Lemma 7 (i). This

Lemma requires bounds for integration with respect to the weight function for δ for various approximation terms, which are provided by Lemma 6. Very similar arguments are also at the core of the proof of part (ii) of Theorem 1.

Proof of Theorem 1:

(iii) We focus on the claim for a flat weighting on θ, the claim for a weighting w on θ follows very similarly.

Let $\hat{f}_T(\theta,\delta)$ be the density of the observations in the pseudo model (13), so that $\widehat{LR}_T(u,\delta) = \hat{f}_T(\theta_0+u,\delta)/\hat{f}_T(\theta_0,0)$. The total variation distance between the posterior distributions computed from the true model density f_T and the pseudo model density $\hat{f}_T$ is then given by

$$\int E_\delta \left| \frac{w(\theta_0+T^{-1/2}u)LR_T(u,\delta)}{m_T} - \frac{w(\theta_0)\widehat{LR}_T(u,\delta)}{\hat{m}_T} \right| du$$
$$\leq \hat{m}_T^{-1} \int E_\delta \left| w(\theta_0+T^{-1/2}u)LR_T(u,\delta) - w(\theta_0)\widehat{LR}_T(u,\delta) \right| du + \hat{m}_T^{-1}|m_T - \hat{m}_T|$$

where $m_T = \int E_\delta w(\theta_0+T^{-1/2}u)LR_T(u,\delta)du > 0$ a.s. and $\hat{m}_T = w(\theta_0)\int E_\delta \widehat{LR}_T(u,\delta)du > 0$ a.s. Since

$$|\hat{m}_T - m_T| \leq \int E_\delta \left| w(\theta_0+T^{-1/2}u)LR_T(u,\delta) - w(\theta_0)\widehat{LR}_T(u,\delta) \right| du \tag{18}$$

it suffices to show that $\int E_\delta \left| w(\theta_0+T^{-1/2}u)LR_T(u,\delta) - w(\theta_0)\widehat{LR}_T(u,\delta) \right| du \xrightarrow{p} 0$ and $\hat{m}_T^{-1} = O_p(1)$.

Now by Fubini's theorem and a direct calculation

$$\int E_\delta \widehat{LR}_T(u,\delta)du = (2\pi)^{k/2}|\tilde{H}|^{-1/2}\exp[\tfrac{1}{2}\hat{u}'\tilde{H}\hat{u}]E_\delta\overline{LR}_T(\delta). \tag{19}$$

Lemma 2 (iii) shows $\hat{u} = O_p(1)$, so that also $\exp[-\frac{1}{2}\hat{u}'\tilde{H}\hat{u}] = O_p(1)$. By Lemma 8, $E_\delta\overline{LR}_T(\delta) \geq 0$ has an absolutely continuous limiting distribution, so that by the continuous mapping theorem, $(E_\delta\overline{LR}_T(\delta))^{-1} = O_p(1)$, and $\hat{m}_T^{-1} = O_p(1)$ follows.

Furthermore, with $\mathcal{S}_T$ and $\mathcal{A}_T$ as defined in Lemma 3,

$$\int E_\delta \left| w(\theta_0+T^{-1/2}u)LR_T(u,\delta) - w(\theta_0)\widehat{LR}_T(u,\delta) \right| du$$
$$\leq \int E_\delta \left| \mathcal{A}_T\mathcal{S}_T w(\theta_0+T^{-1/2}u)LR_T(u,\delta) - w(\theta_0)\widehat{LR}_T(u,\delta) \right| du$$
$$+ \int E_\delta(1-\mathcal{A}_T\mathcal{S}_T)w(\theta_0+T^{-1/2}u)LR_T(u,\delta)du.$$

The last term converges in probability to zero by Lemma 3, part (i). Also

$$\int E_\delta \left| \mathcal{A}_T\mathcal{S}_T w(\theta_0+T^{-1/2}u)LR_T(u,\delta) - w(\theta_0)\widehat{LR}_T(u,\delta) \right| du$$
$$\leq \int |w(\theta_0+T^{-1/2}u) - w(\theta_0)|E_\delta\mathcal{A}_T\mathcal{S}_T LR_T(u,\delta)du + w(\theta_0)\int E_\delta \left| \mathcal{A}_T\mathcal{S}_T LR_T(u,\delta) - \widehat{LR}_T(u,\delta) \right| du.$$

The last term converges in probability to zero by Lemma 7, part (i). For the first term after the inequality, we compute

$$\int |w(\theta_0 + T^{-1/2}u) - w(\theta_0)| E_\delta \mathcal{A}_T \mathcal{S}_T LR_T(u,\delta) du$$
$$\leq \sup_{||u||<a_T} |w(\theta_0 + T^{-1/2}u) - w(\theta_0)| \left(\int E_\delta |\mathcal{A}_T \mathcal{S}_T LR_T(u,\delta) - \widehat{LR}_T(u,\delta)| du + w(\theta_0)^{-1} \hat{m}_T \right).$$

But $T^{-1/2}a_T \to 0$ and the continuity of w at θ_0 imply $\sup_{||u||<a_T} |w(\theta_0 + T^{-1/2}u) - w(\theta_0)| \to 0$. Furthermore, as shown above, $\hat{m}_T = O_p(1)$, and the result follows from Lemma 7 (i).

The convergence in probability under the unstable model follows from Lemma 1.

(i) For brevity, we again focus on the case of a flat weighting on θ only.

By definition of the weighted average risk and Fubini's Theorem

$$\begin{aligned}
& WAR(\hat{a}) \\
= & \int w(\theta_0) E_\delta \int L_T(\theta_0,\delta,\hat{a}) f_T(\theta_0,\delta) d\mu_T d\theta_0 \\
= & \int \frac{\int E_\delta L_T(\theta,\delta,\hat{a}) f_T(\theta,\delta) w(\theta) d\theta}{\int E_\delta f_T(\theta,\delta) w(\theta) d\theta} \int E_\delta f_T(\theta_0,\delta) w(\theta_0) d\theta_0 d\mu_T \\
= & \int w(\theta_0) \int \frac{\int E_\delta L_T(\theta_0 + T^{-1/2}u,\delta,\hat{a}) LR_T(u,\delta) w(\theta_0 + T^{-1/2}u) du}{m_T} E_\delta f_T(\theta_0,\delta) d\mu_T d\theta_0.
\end{aligned}$$

Similarly, define

$$\widehat{WAR}(\hat{a}) = \int w(\theta_0) \int \frac{\int E_\delta w(\theta_0) \widehat{LR}_T(u,\delta) L_T(\theta_0 + T^{-1/2}u,\delta,\hat{a}) du}{\hat{m}_T} E_\delta f_T(\theta_0,\delta) d\mu_T d\theta_0. \tag{20}$$

Note that

$$\begin{aligned}
& \sup_{a \in \mathbb{A}_T} |WAR(a) - \widehat{WAR}(a)| \\
\leq & \bar{L} \int w(\theta_0) \int \int E_\delta \left| \frac{LR_T(u,\delta) w(\theta_0 + T^{-1/2}u)}{m_T} - \frac{w(\theta_0) \widehat{LR}_T(u,\delta)}{\hat{m}_T} \right| du E_\delta f_T(\theta_0,\delta) d\mu_T d\theta_0.
\end{aligned}$$

Now since $m_T > 0$ and $\hat{m}_T > 0$ a.s., we have

$$\int E_\delta \left| \frac{LR_T(u,\delta) w(\theta_0 + T^{-1/2}u)}{m_T} - \frac{w(\theta_0) \widehat{LR}_T(u,\delta)}{\hat{m}_T} \right| du$$
$$\leq \int E_\delta \left(m_T^{-1} LR_T(u,\delta) w(\theta_0 + T^{-1/2}u) + \hat{m}_T^{-1} w(\theta_0) \widehat{LR}_T(u,\delta) \right) du = 2$$

almost surely. Let $M_T = E_\delta \prod_{t=1}^T \mathbf{1}[(\theta_0 + \delta_t) \in \Theta] > 0$. Since Θ contains an open ball around θ_0 and $\sup_{\lambda \in [0,1]} ||G(\lambda)||$ is bounded almost surely, $M_T \to 1$. Note that for all T, $M_T^{-1} E_\delta f_T(\theta_0,\delta)$ is a probability density with respect to μ_T, so that the convergence in probability

$$\int E_\delta \left| \frac{LR_T(u,\delta) w(\theta_0 + T^{-1/2}u)}{m_T} - \frac{w(\theta_0) \widehat{LR}_T(u,\delta)}{\hat{m}_T} \right| du \xrightarrow{p} 0$$

established in part (i) of this proof under the unstable model with density $M_T^{-1}E_\delta f_T(\theta_0,\delta)$ implies via dominated convergence that

$$M_T \int\int E_\delta \left| \frac{LR_T(u,\delta)w(\theta_0+T^{-1/2}u)}{m_T} - \frac{w(\theta_0)\widehat{LR}_T(u,\delta)}{\hat{m}_T} \right| du M_T^{-1}E_\delta f_T(\theta_0,\delta)d\mu_T \to 0$$

for almost all θ_0. Since this is also bounded by 2, by another application of the dominated convergence theorem, we have

$$\int w(\theta_0)\int\int E_\delta \left| \frac{LR_T(u,\delta)w(\theta_0+T^{-1/2}u)}{m_T} - \frac{w(\theta_0)\widehat{LR}_T(u,\delta)}{\hat{m}_T} \right| du E_\delta f_T(\theta_0,\delta)d\mu_T d\theta_0 \to 0.$$

Since for any $\hat{a}$, $\widehat{WAR}(\hat{a})-\widehat{WAR}(\hat{a}^*)\geq 0$ by the definition of $\hat{a}^*$ and $\widehat{WAR}(\hat{a})$,

$$\begin{aligned} WAR(\hat{a})-WAR(\hat{a}^*) &= \left(\widehat{WAR}(\hat{a})-\widehat{WAR}(\hat{a}^*)\right) \\ &\quad +\left(WAR(\hat{a})-\widehat{WAR}(\hat{a})\right)+\left(\widehat{WAR}(\hat{a}^*)-WAR(\hat{a}^*)\right) \\ &\geq \left(WAR(\hat{a})-\widehat{WAR}(\hat{a})\right)+\left(\widehat{WAR}(\hat{a}^*)-WAR(\hat{a}^*)\right)\to 0. \end{aligned}$$

(ii) By the Neyman Pearson Lemma and Fubini's Theorem, the weighted average power maximizing test of (9) under Condition 2 weighting rejects for large values of $E_\delta LR_T(0,\delta-e\bar{\delta})$, and the weighted average power maximizing test in the pseudo model (14) rejects for large values of $E_\delta\overline{LR}_T(\delta)$. We have

$$\begin{aligned} |E_\delta LR_T(0,\delta-e\bar{\delta})-E_\delta\overline{LR}_T(\delta)| &\leq E_\delta|\mathcal{S}_T LR_T(0,\delta-e\bar{\delta})-\overline{LR}_T(\delta)| \\ &\quad +E_\delta(1-\mathcal{S}_T)LR_T(0,\delta-e\bar{\delta})\xrightarrow{p}0 \end{aligned}$$

by applying Lemmas 3 (ii) and 7 (ii). Furthermore, the asymptotic distribution of $E_\delta\overline{LR}_T(\delta)$ under the null hypothesis is absolutely continuous by Lemma 8, so that the result follows from the second claim in Lemma 1 by the same arguments as employed in Andrews and Ploberger (1994) in the proof of their Theorem 2.

Proof of Theorem 2:

(iii) In matrix form, the pseudo model (13) is $\hat{s}+D_{\tilde{h}}e\hat{\theta}|(D_{\tilde{h}},\delta,\theta)\sim\mathcal{N}(D_{\tilde{h}}(\delta+e\theta),D_{\tilde{h}})$, so that conditionally on $D_{\tilde{h}}$ and θ only,

$$\begin{pmatrix} \hat{s}+D_{\tilde{h}}e\hat{\theta} \\ \delta \end{pmatrix}|(D_{\tilde{h}},\theta)\sim\mathcal{N}\left(\begin{pmatrix} D_{\tilde{h}}e\theta \\ 0\end{pmatrix},\begin{pmatrix} D_{\tilde{h}}+D_{\tilde{h}}\Sigma_\delta D_{\tilde{h}} & D_{\tilde{h}}\Sigma_\delta \\ \Sigma_\delta D_{\tilde{h}} & \Sigma_\delta\end{pmatrix}\right).$$

Using the identity

$$(I_{Tk}+D_{\tilde{h}}\Sigma_\delta)^{-1}=I_{Tk}-(I_{Tk}+D_{\tilde{h}}\Sigma_\delta)^{-1}D_{\tilde{h}}\Sigma_\delta \qquad (21)$$

we find with $K=\Sigma_\delta D_{\tilde{h}}(D_{\tilde{h}}+D_{\tilde{h}}\Sigma_\delta D_{\tilde{h}})^{-1}=\Sigma_\delta-\Sigma_\delta D_{\tilde{h}}(D_{\tilde{h}}+D_{\tilde{h}}\Sigma_\delta D_{\tilde{h}})^{-1}D_{\tilde{h}}\Sigma_\delta$ that

$$\delta|(\hat{s}+D_{\tilde{h}}e\hat{\theta},D_{\tilde{h}},\theta)\sim\mathcal{N}(K(\hat{s}+D_{\tilde{h}}e(\hat{\theta}-\theta)),K).$$

Furthermore, with a flat prior, the posterior for θ is proportional to the likelihood, so that $(\hat{s}+D_{\tilde{h}}e\hat{\theta})|(D_{\tilde{h}},\theta)\sim\mathcal{N}(D_{\tilde{h}}e\theta,D_{\tilde{h}}+D_{\tilde{h}}\Sigma_\delta D_{\tilde{h}})$ implies $\theta|(\hat{s}+D_{\tilde{h}}e\hat{\theta},D_{\tilde{h}})\sim\mathcal{N}((e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}D_{\tilde{h}}^{-1}\hat{s}+\hat{\theta},(e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1})$. Thus

$$\begin{aligned}\begin{pmatrix}\delta\\ \theta\end{pmatrix}|(\hat{s}+D_{\tilde{h}}e\hat{\theta},D_{\tilde{h}}) &\sim \mathcal{N}\left(\begin{pmatrix}K(\hat{s}-D_{\tilde{h}}e(e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}D_{\tilde{h}}^{-1}\hat{s})\\ (e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}D_{\tilde{h}}^{-1}\hat{s}+\hat{\theta}\end{pmatrix},V_{\delta\theta}\right)\\ \text{where } V_{\delta\theta} &= \begin{pmatrix}K+KD_{\tilde{h}}e(e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}e'D_{\tilde{h}}K & KD_{\tilde{h}}e(e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}\\ (e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}e'D_{\tilde{h}}K & (e'(D_{\tilde{h}}^{-1}+\Sigma_\delta)^{-1}e)^{-1}\end{pmatrix}\end{aligned}$$

and employing once more (21), we conclude $\delta+e\theta|(\hat{s}+D_{\tilde{h}}e\hat{\theta},D_{\tilde{h}})\sim\mathcal{N}(e\hat{\theta}+\Sigma\hat{s},\Sigma)$.

(i) Immediate from Theorem 1 (i) and the proof of part (i).

(ii) Let $\bar{R}(\delta)=\exp[-\frac{1}{2}\delta'D_{\tilde{h}}\delta+\frac{1}{2}\delta'D_{\tilde{h}}e(e'D_{\tilde{h}}^{-1}e)^{-1}e'D_{\tilde{h}}\delta]$. Using (19), we find

$$\frac{E_\delta\overline{LR}_T(\delta)}{E_\delta\bar{R}(\delta)}=\exp[\tfrac{1}{2}\hat{s}'\Sigma\hat{s}].$$

By Lemma 7 (iv), $E_\delta\bar{R}(\delta)-E_\delta\exp[-\frac{1}{2}\delta'D_\Gamma(\delta-e\bar{\delta})]\xrightarrow{p}0$. By the CMT, $\exp[-\frac{1}{2}\delta'D_\Gamma(\delta-e\bar{\delta})]\Rightarrow\exp[-\frac{1}{2}\int G^*\Gamma G^*]$, and since $\bar{R}(\delta)<1$ a.s., also $E_\delta\bar{R}(\delta)\to E_G\exp[-\frac{1}{2}\int G^{*\prime}\Gamma G^*]$. The result now follows from Lemma 8.

Proof of Theorem 3:

We write $\hat{\Lambda}_t$ for $\hat{\Lambda}_{T,t}$ to enhance readability.

For the first claim, note that if $\{\tilde{h}_t\}_{t=1}^T$ satisfies (12) under the stable model, so does $\check{h}_t=\hat{\Lambda}_t\tilde{h}_t\hat{\Lambda}_t'$. Thus, with $D_{\hat{\Lambda}}=\mathrm{diag}(\hat{\Lambda}_1,\cdots,\hat{\Lambda}_T)$ and $D_{\check{h}}=\mathrm{diag}(\check{h}_1,\cdots,\check{h}_T)$,

$$E_\delta\overline{LR}_T(D_{\hat{\Lambda}}\delta)=E_\delta\exp[\hat{s}'D_{\hat{\Lambda}}\delta-\tfrac{1}{2}\delta'D_{\check{h}}\delta+\tfrac{1}{2}\delta'D_{\check{h}}e(e'D_{\check{h}}e)^{-1}e'D_{\check{h}}\delta]$$

By summation by parts with $\hat{\Lambda}_0=\hat{\Lambda}_1$

$$T^{-1/2}\sum_{j=1}^t\hat{\Lambda}_j's_j(\hat{\theta})=\hat{\Lambda}_t'T^{-1/2}\sum_{j=1}^t s_j(\hat{\theta})-\sum_{j=1}^t(\hat{\Lambda}_j'-\hat{\Lambda}_{j-1}')(T^{-1/2}\sum_{j=1}^{t-1}s_j(\hat{\theta}))$$

so that by Lemma 2 (iv) and the assumptions on $\hat{\Lambda}_t$, $\sup_{t\le T}||T^{-1/2}\sum_{j=1}^t(\hat{\Lambda}_j'-I_k)s_j(\hat{\theta})||\le(\sup_{t\le T}||T^{-1/2}\sum_{j=1}^t s_j(\hat{\theta})||)(\sum_{t=1}^T||\hat{\Lambda}_t-\hat{\Lambda}_{t-1}||+\sup_{t\le T}||\hat{\Lambda}_t-I_k||)\xrightarrow{p}0$. Proceeding as in the proof of Theorem 1 (ii), it is seen that the only additional complication arises through the additional term $\sum s_t(\hat{\theta})'(\hat{\Lambda}_t-I_k)\delta_t$ in the definition of ς_T^* and ς_T^0 (and thus $\bar{\varsigma}_T^*$, $\underline{\varsigma}_T^*$, $\bar{\varsigma}_T^0$ and $\underline{\varsigma}_T^0$) in Lemma 7 (ii) and (iii). By letting $\xi_t=(\hat{\Lambda}_t'-I_k)s_t(\hat{\theta})$ while invoking Lemma 6 (ii), it continues to be the case that

$E_\delta \exp 2\bar{\varsigma}_T^* \xrightarrow{p} 1$, $E_\delta \exp \underline{\varsigma}_T^* \xrightarrow{p} 1$, $E_\delta \exp 2\bar{\varsigma}_T^0 \xrightarrow{p} 1$ and $E_\delta \exp \underline{\varsigma}_T^0 \xrightarrow{p} 1$ even under this new definition, so the result follows as before.

For the second claim, proceed as in the proof of part (ii) of Theorem 1, but with $\widehat{WAR}(\hat{a})$ in (20) substituted by

$$\widehat{WAR}_\Lambda(\hat{a}) = \int w(\theta_0) \int \frac{\int E_\delta w(\theta_0)\widehat{LR}_T(u, D_{\hat{\Lambda}}\delta)L(\theta_0 + T^{-1/2}u, D_{\hat{\Lambda}}\delta, \hat{a})du}{\hat{m}_{\Lambda,T}} E_\delta f_T(\theta_0, \delta)d\mu_T d\theta_0$$

where $\hat{m}_{\Lambda,T} = \int E_\delta w(\theta_0)\widehat{LR}_T(u, D_{\hat{\Lambda}}\delta)du$. We have

$$\begin{aligned} & |\widehat{WAR}_\Lambda(\hat{a}) - \int w(\theta_0) \int \frac{\int E_\delta w(\theta_0)\widehat{LR}_T(u, D_{\hat{\Lambda}}\delta)L(\theta_0 + T^{-1/2}u, \delta, \hat{a})du}{\hat{m}_{\Lambda,T}} E_\delta f_T(\theta_0, \delta)d\mu_T d\theta_0| \\ \leq \; & \int w(\theta_0) \left(\sup_{\theta\in\Theta,\delta\in\mathbb{R}^{Tk},a\in\mathbb{A}_T} |L_T(\theta, D_{\hat{\Lambda}}\delta, a) - L_T(\theta, \delta, a)| \right) E_\delta f_T(\theta_0, \delta)d\mu_T d\theta_0 \to 0 \end{aligned}$$

where the convergence follows from $\sup_{\theta\in\Theta,\delta\in\mathbb{R}^{Tk},a\in\mathbb{A}_T} |L_T(\theta, D_{\hat{\Lambda}}\delta, a) - L_T(\theta, \delta, a)| \xrightarrow{p} 0$ in the stable model, Lemma 1 and the dominated convergence theorem as $\sup_{\theta\in\Theta,\delta\in\mathbb{R}^{Tk},a\in\mathbb{A}_T} |L_T(\theta, D_{\hat{\Lambda}}\delta, a) - L_T(\theta, \delta, a)| \leq 2\bar{L}$. It thus suffices to proceed as in the proof of Theorem 1 with $\widehat{LR}_T(u, \delta)$ replaced by

$$\begin{aligned} & \widehat{LR}_T(u, D_{\hat{\Lambda}}\delta) \\ = \; & \exp(\hat{s}'D_{\hat{\Lambda}}\delta - \tfrac{1}{2}\delta' D_{\tilde{h}}\delta + T^{-1/2}(\hat{u} - u)'e'D_{\tilde{h}}D_{\hat{\Lambda}}\delta - \tfrac{1}{2}T^{-1}u'e(e'D_{\tilde{h}}e)^{-1}u'e + T^{-1}\hat{u}'e(e'D_{\tilde{h}}e)^{-1}u'e). \end{aligned}$$

The difference between $\hat{s}'D_{\hat{\Lambda}}\delta$ and $\hat{s}'\delta$ can again be handled by suitably modifying ς_T in Lemma 7 (i) as in the proof of the first claim, and further inspection of the proof reveals that the only important properties of the $Tk \times Tk$ matrices in the innerproducts with respect to (δ, δ), (e, e) and (e, δ) are that (i) they are block diagonal constructed from $k \times k$ matrices satisfying (12) in the stable model, and (ii) that the ones for (δ, δ) and (e, e) are symmetric and positive definite, both of which remains the case for $\widehat{LR}_T(u, D_{\hat{\Lambda}}\delta)$.

Proof of Theorem 4:

For the claim regarding the analogous statement of Theorem 1 (iii), proceed up to equation (18) as in the proof of Theorem 1 (iii) with E_δ now denoting integration with respect to the mixture. With $E_{\delta(i)}$ denoting integration with respect to the measure of $\{T^{-1/2}G_i(t/T)\}$, it then suffices to show that $\int E_{\delta(i)} \left| w(\theta_0 + T^{-1/2}u)LR_T(u, \delta_{(i)}) - w(\theta_0)\widehat{LR}_T(u, \delta_{(i)}) \right| du \xrightarrow{p} 0$ for $i = 1, \cdots, n_G$ and $\hat{m}_T^{*-1} = (w(\theta_0)\sum_i p_i \int E_{\delta(i)}\widehat{LR}_T(u, \delta_{(i)})du)^{-1} = O_p(1)$. Without loss of generality, assume $p_1 > 0$. Then $\hat{m}_T^*/w(\theta_0)p_1 \geq \int E_{\delta(1)}\widehat{LR}_T(u, \delta_{(1)})du$, and the result $\hat{m}_T^{*-1}$ follows from the same reasoning as in the proof of Theorem 1 (iii). The result now follows by proceeding as in the remainder of the proof of Theorem 1 (iii) and by invoking invoking Lemmas 3 (i) and 7 (i) for each of the n_G components in the measure of δ.

The claim regarding the analogous statement of Theorem 1 (i) follows as in the proof of Theorem 1 (i) from this result by substituting integrations with respect to δ by integrations with respect to the mixture.

For the claim regarding Theorem 2 and the mixing probabilities, note that

$$\begin{aligned} & E_{\delta(i)} \int (2\pi)^{-k/2} |\tilde{H}|^{1/2} \exp[-\tfrac{1}{2}\hat{u}'\tilde{H}\hat{u}]\widehat{LR}_T(u, \delta_{(i)})du \\ = \;& E_{\delta(i)} \int (2\pi)^{-k/2} |\tilde{H}|^{1/2} \exp[\hat{s}'\delta_{(i)} - \tfrac{1}{2}\delta_{(i)}' D_{\tilde{h}} \delta_{(i)} + T^{-1/2}(\hat{u}-u)'e'D_{\tilde{h}}\delta_{(i)} - \tfrac{1}{2}(u-\hat{u})'\tilde{H}(u-\hat{u})]du \\ = \;& E_{\delta(i)} \overline{LR}_T(\delta_{(i)}) \\ = \;& |D_{\tilde{h}}\Sigma_\delta + I_{Tk}|^{-1/2} |e'D_{\tilde{h}}e - e'D_{\tilde{h}}K_i D_{\tilde{h}}e|^{-1/2} \exp[\tfrac{1}{2}\hat{s}'\Sigma_i \hat{s}] \end{aligned}$$

so that the posterior odds of model i and model j in the pseudo model are as claimed.

5.2.3 Additional Lemmas

Lemma 2 *Under Condition 1:*

(i) $T^{-1/2}\sum_{t=1}^{[\cdot T]} s_t(\theta_0) \Rightarrow \int_0^{\cdot} \Gamma^{1/2}(l)dW(l)$, *where W is a $k\times 1$ standard Wiener process*

(ii) $\sup_{t\leq T, \{v_t\}_{t=1}^T \in \mathcal{C}_T^T, \{\tilde{v}_t\}_{t=1}^T\in\mathcal{C}_T^T} T^{-1}||\sum_{s=1}^t (2\int_0^1 \lambda h_s(\theta_0+v_s+\lambda\tilde{v}_s)d\lambda - \Gamma_s)|| \xrightarrow{p} 0$ *and* $\sup_{t\leq T, \{v_t\}_{t=1}^T \in \mathcal{C}_T^T, \{\tilde{v}_t\}_{t=1}^T\in\mathcal{C}_T^T} T^{-1}||\sum_{s=1}^t (\int_0^1 h_s(\theta_0+\lambda(v_s-\tilde{v}_s))d\lambda - \Gamma_s)|| \xrightarrow{p} 0$, *where $\mathcal{C}_T$ is an arbitrary decreasing neighborhood of θ_0, and $\mathcal{C}_T^T = \mathcal{C}_T\times\cdots\times\mathcal{C}_T$*

(iii) $\hat{u} = T^{1/2}(\hat{\theta}-\theta_0) = O_p(1)$

(iv) $\sup_{t\leq T} ||T^{-1/2}\sum_{s=t}^T s_s(\hat{\theta})|| = O_p(1)$

(v) $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]} s_t(\hat{\theta})s_t(\hat{\theta})' - \int_0^\lambda \Gamma(l)dl|| \xrightarrow{p} 0$ *and* $T^{-1}\sum s_t(\theta_0)s_t(\theta_0)' = O_p(1)$

(vi) $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]} h_t(\hat{\theta}) - \int_0^\lambda \Gamma(l)dl|| \xrightarrow{p} 0$

Proof. (i) Fix any $k\times 1$ vector v with $v'v=1$, and let $\eta_t = v's_t(\theta_0)$. Then $\{\eta_t, \mathfrak{F}_t\}$ is a martingale difference array and $T^{-1}\sum_{t=1}^T E[|\eta_t|^{2+\varepsilon}|\mathfrak{F}_{t-1}] \leq T^{-1}\sum_{t=1}^T E[||s_t(\theta_0)||^{2+\varepsilon}|\mathfrak{F}_{t-1}] = O_p(1)$ by Condition 1 (MDA). Let $\omega_\eta^2 = \int_0^1 v'\Gamma(l)vdl$ and $g(\lambda) = \int_0^\lambda v'\Gamma(l)vdl/\omega_\eta^2$, which is a continuous and strictly increasing function on the unit interval, so that it has an inverse g^{-1}. By Corollary 3.8 of McLeish (1974), $T^{-1/2}\sum_{t=1}^{[g^{-1}(\cdot)T]}\eta_t \Rightarrow \omega_\eta W_\eta(\cdot)$, where W_η is a standard scalar Wiener process and the convergence is on the space of cadlag functions on the unit interval, equipped with the Skorohod norm. By the continuous mapping theorem, we hence obtain $T^{-1/2}\sum_{t=1}^{[\cdot T]}\eta_t \Rightarrow \omega_\eta W_\eta(g(\cdot)) \sim v'\int_0^\cdot \Gamma(l)^{1/2}dW(l)$ and the result follows from the Functional Cramer-Wold device (see, for instance, Proposition 7.26 of White (2001)).

(ii) We have

$$\begin{aligned} & T^{-1}||\sum_{s=1}^t (2\int_0^1 \lambda h_s(\theta_0+v_s+\lambda\tilde{v}_s)d\lambda - \Gamma_s)|| \\ & \qquad \leq T^{-1}||\sum_{s=1}^t (2\int_0^1 \lambda h_s(\theta_0+v_s+\lambda\tilde{v}_s)d\lambda - h_s(\theta_0))|| + T^{-1}||\sum_{s=1}^t (\Gamma_s - h_s(\theta_0))||. \end{aligned}$$

Now $\sup_{t\leq T} T^{-1}||\sum_{s=1}^{t}(\Gamma_s - h_s(\theta_0))|| \xrightarrow{p} 0$ by Condition 1 (LLLN) and $\sup_{\lambda\in[0,1]}||T^{-1}\sum_{s=1}^{[\lambda T]}\Gamma_s - \int_0^\lambda \Gamma(s)ds|| \to 0$, and

$$\sup_{t\leq T,\{v_t\}_{t=1}^T\in\mathcal{C}_T^T,\{\tilde{v}_t\}_{t=1}^T\in\mathcal{C}_T^T} T^{-1}||\sum_{s=1}^{t} 2\int_0^1 \lambda(h_s(\theta_0+v_s+\lambda\tilde{v}_s)-h_s(\theta_0))d\lambda||$$

$$\leq 2T^{-1}\sum_{t=1}^{T}\sup_{v\in\mathcal{C}_T}||h_t(\theta_0+2v)-h_t(\theta_0))|| \xrightarrow{p} 0$$

by Condition 1 (LLLN). The second claim follows similarly.

(iii) For any $\varepsilon > 0$,

$$\begin{aligned} P(||\hat{\theta}-\theta_0|| &\geq \varepsilon) \leq P(\sup_{||\theta-\theta_0||\geq\varepsilon} T^{-1}\sum[l_t(\theta)-l_t(\theta_0)] \geq -K(\varepsilon)) \\ &\leq 1 - P(\sup_{||\theta-\theta_0||\geq\varepsilon} T^{-1}\sum \sup_{||v||<T^{-1/2+\eta}}[l_t(\theta+v)-l_t(\theta_0)] < -K(\varepsilon)) \to 0 \end{aligned}$$

by Condition 1 (ID) and so $\hat{\theta} \xrightarrow{p} \theta_0$.

Further, as $\hat{\theta} \xrightarrow{p} \theta_0$, there exists a sequence of decreasing $\mathcal{T}_T$ neighborhoods of θ_0 such that $P(\hat{\theta}\in\mathcal{T}_T) \to 1$. For $v \in \Theta_0$, we have by the fundamental theorem of calculus applied row by row that $s_t(\theta_0+v) - s_t(\theta_0) = \left(-\int_0^1 h_t(\theta_0+\lambda v)d\lambda\right)v$ almost surely for $t = 1,\cdots,T$. Let T be large enough so that $\mathcal{T}_T \subset \Theta_0$, and define $h_t^S = \int_0^1 h_t(\theta_0+\lambda(\hat{\theta}-\theta_0))d\lambda$ if $\hat{\theta}\in\mathcal{T}_T$, and $h_t^S = \tilde{h}_t$ otherwise, so that from the first order condition $\mathbf{1}[\hat{\theta}\in\mathcal{T}_T]\sum s_t(\hat{\theta}) = 0$, we obtain

$$\mathbf{1}[\hat{\theta}\in\mathcal{T}_T]\left(T^{-1/2}\sum s_t(\theta_0) - \left(T^{-1}\sum h_t^S\right)T^{1/2}(\hat{\theta}-\theta_0)\right) = 0 \tag{22}$$

almost surely for $t = 1,\cdots,T$. From part (i), $T^{-1/2}\sum s_t(\theta_0) = O_p(1)$. Applying the result of part (ii), we obtain $T^{-1}\sum h_t^S - T^{-1}\sum\Gamma_t \xrightarrow{p} 0$. But $T^{-1}\sum\Gamma_t \to \int\Gamma$, which is positive definite, so the result follows from (22) and $P(\hat{\theta}\in\mathcal{T}_T) \to 1$.

(iv) Proceed as in the proof of part (iii) to obtain

$$\mathbf{1}[\hat{\theta}\in\mathcal{T}_T]\left(T^{-1/2}\sum_{s=1}^{t}s_s(\hat{\theta}) - T^{-1/2}\sum_{s=1}^{t}s_s(\theta_0) + \left(T^{-1}\sum_{s=1}^{t}h_s^S\right)T^{1/2}(\hat{\theta}-\theta_0)\right) = 0$$

almost surely, so that

$$\sup_{t\leq T}||T^{-1/2}\sum_{s=t}^{T}s_s(\hat{\theta})|| \leq \sup_{t\leq T}||T^{-1/2}\sum_{s=t}^{T}s_s(\theta_0)|| + T^{1/2}\sup_{t\leq T}||T^{-1}\sum_{s=t}^{T}h_s^S||\cdot||\hat{\theta}-\theta_0|| + o_p(1)$$

and the result follows from parts (i), (ii) and (iii) of this Lemma and the CMT.

(v) From the proof of part (iii), $\mathbf{1}[\hat{\theta}\in\mathcal{T}_T](s_t(\hat{\theta}) - s_t(\theta_0) + h_s^S(\hat{\theta}-\theta_0)) = 0$, almost surely for $t = 1,\cdots,T$, so that

$$\sup_{\lambda\in[0,1]}||T^{-1}\sum_{t=1}^{[\lambda T]}s_t(\hat{\theta})s_t(\hat{\theta})' - T^{-1}\sum_{t=1}^{[\lambda T]}s_t(\theta_0)s_t(\theta_0)'||$$

$$\leq 2||\hat{u}||T^{-1}\sum||h_t^S||T^{-1/2}\sup_{t\leq T}||s_t(\theta_0)|| + T^{-1}||\hat{u}||^2\sup_{t\leq T}||h_t^S||T^{-1}\sum||h_t^S||$$

with probability $P(\hat{\theta} \in \mathcal{T}_T) \to 1$. Now $T^{-1}\sum ||h_t^S|| = O_p(1)$ and $||\hat{u}|| = O_p(1)$ from parts (ii) and (iii) and Condition 1 (LLLN), and $T^{-1/2}\sum_{t=1}^{[\cdot T]} s_t(\theta_0) \Rightarrow \int_0^{\cdot} \Gamma^{1/2}(l)dW(l)$ implies $T^{-1/2}\sup_{t\le T}||s_t(\theta_0)|| \xrightarrow{p} 0$, and also $T^{-1}\sup_{t\le T}||h_t^S|| \le T^{-1}\sup_{t\le T}||h_t(\theta_0)|| + T^{-1}\sup_{t\le T,\theta\in\mathcal{T}_T}||h_t(\theta_0) - h_t(\theta)|| \xrightarrow{p} 0$ by Condition 1 (LLLN) (i) and (iii), so that $\sup_{\lambda\in[0,1]}||T^{-1}\sum_{t=1}^{[\lambda T]} s_t(\hat{\theta})s_t(\hat{\theta})' - T^{-1}\sum_{t=1}^{[\lambda T]} s_t(\theta_0)s_t(\theta_0)'|| \xrightarrow{p} 0$.

Let $v \in \mathbb{R}^k$, and define $\eta_t = v's_t(\theta_0)$. Then from Condition 1 (MDA), $\{\eta_t, \mathfrak{F}_t\}$ is a martingale difference array with conditional variance process $V_{\eta,t}^2 = v'E[s_t(\theta_0)s_t(\theta_0)'|\mathfrak{F}_{t-1}]v$, and $\sup_{\lambda\in[0,1]}|T^{-1}\sum_{t=1}^{[\lambda T]} V_{\eta,t}^2 - v'(\int_0^\lambda \Gamma(l)dl)v| \xrightarrow{p} 0$. Note that $T^{-1}\sum_{t=1}^T V_{\eta,t}^2 \le ||v||^2 T^{-1}\sum_{t=1}^T E[||s_t(\theta_0)||^{2+\epsilon}|\mathfrak{F}_{t-1}] = O_p(1)$ implies $T^{-1}\sum E[\eta_t^2 \mathbf{1}[|\eta_t| > T^{1/2}c]|\mathfrak{F}_{t-1}] \xrightarrow{p} 0$ for all $0 < c < \infty$, so that from Theorem 2.23 of Hall and Heyde (1980), $\sup_{t\le T}|T^{-1}\sum_{s=1}^t(\eta_s^2 - V_{\eta,s}^2)| = \sup_{\lambda\in[0,1]}|T^{-1}v'\sum_{t=1}^{[\lambda T]}(s_t(\theta_0)s_t(\theta_0)' - E[s_t(\theta_0)s_t(\theta_0)'|\mathfrak{F}_{t-1}])v| \xrightarrow{p} 0$, so that also $\sup_{\lambda\in[0,1]}|T^{-1}v'\sum_{t=1}^{[\lambda T]}(s_t(\theta_0)s_t(\theta_0)' - \int_0^\lambda \Gamma(l)dl)v| \xrightarrow{p} 0$. This holds for arbitrary $v \in \mathbb{R}^k$, so in particular jointly for all vectors v_j, $j = 1, \cdots, 2^k$ with elements that are either zero or one. It is easy to see that if $v_j'A_0v_j = v_j'A_1v_j$ for all such v_j, $j = 1, \cdots, 2^k$ for two symmetric matrices A_0 and A_1, then $A_0 = A_1$, and both results follow.

(vi) Follows from parts (ii) and (iii). ■

Lemma 3 *Under Conditions 1 and 2, there exists a sequence of real numbers a_T with $a_T \to \infty$ and $T^{-1/2}a_T \to 0$ such that*
(i) $\int w(\theta_0 + T^{-1/2}u)E_\delta(1 - \mathcal{A}_T\mathcal{S}_T)LR_T(u,\delta)du \xrightarrow{p} 0$
(ii) $E_\delta(1 - \mathcal{S}_T)LR_T(0, \delta - e\bar{\delta}) \xrightarrow{p} 0$

Proof. (i) For any choice of a_T, we have

$$\begin{aligned} & |\int w(\theta_0 + T^{-1/2}u)E_\delta(1 - \mathcal{A}_T\mathcal{S}_T)LR_T(u,\delta)du| \\ \le & |\int w(\theta_0 + T^{-1/2}u)E_\delta(1 - \mathcal{S}_T)LR_T(u,\delta)du| + |\int w(\theta_0 + T^{-1/2}u)E_\delta\mathcal{S}_T(1 - \mathcal{A}_T)LR_T(u,\delta)du| \\ = & \rho_1 + \rho_2. \end{aligned}$$

For $\rho_1 = \int w(\theta_0 + T^{-1/2}u)E_\delta(1 - \mathcal{S}_T)LR_T(u,\delta)du$, note that by Markov's inequality, for any $\epsilon > 0$

$$\begin{aligned} P(\rho_1 &> \epsilon) \le \epsilon^{-1}E\rho_1 \\ &= \epsilon^{-1}\int\int w(\theta_0 + T^{-1/2}u)E_\delta(1 - \mathcal{S}_T)LR_T(u,\delta)du f_T(\theta_0, 0)d\mu_T \\ &= \epsilon^{-1}\int w(\theta_0 + T^{-1/2}u)E_\delta(1 - \mathcal{S}_T)\int f_T(\theta_0 + T^{-1/2}u, \delta)d\mu_T du \end{aligned}$$

where the interchange of the order of integration in the second equality follows from Fubini's Theorem. For any fixed u and δ, if for all $t \le T$, $\theta_0 + T^{-1/2}u + \delta_t \in \Theta$, then $f_T(\theta_0 + T^{-1/2}u, \delta)$ is a probability density with respect to μ_T, and $\int f_T(\theta_0 + T^{-1/2}u, \delta)d\mu_T = 1$. If for some t, $\theta_0 + T^{-1/2}u + \delta_t \notin \Theta$, then

$f_T(\theta_0 + T^{-1/2}u, \delta) = 0$, and also $\int f_T(\theta_0 + T^{-1/2}u, \delta) d\mu_T = 0$. Therefore, $\sup_{u \in \mathbb{R}^k, \delta \in \mathbb{R}^{Tk}} \int f_T(\theta_0 + T^{-1/2}u, \delta) d\mu_T \leq 1$, and we obtain

$$\begin{aligned} P(\rho_1 &> \epsilon) \leq \epsilon^{-1} \int w(\theta_0 + T^{-1/2}u) E_\delta (1 - \mathcal{S}_T) du \\ &= \epsilon^{-1} \int w(\theta_0 + T^{-1/2}u) du \ E_\delta \mathbf{1}[T^{1/2} \sup_{t \leq T} ||\delta_t|| > T^\eta]. \end{aligned}$$

Now by a change of variable $\int w(\theta_0 + T^{-1/2}u) du = T^{k/2} \int w(\theta) d\theta = T^{k/2}$. Furthermore, let $\bar{G} = \sup_{t \leq T, T, i \leq k} |G_{(i)}(t/T)|$, where $G_{(i)}(s)$ is the ith element of $G(s)$. Then $\bar{G} \leq \sup_{i, s \in [0,1]} |G_{(i)}(s)|$, which is bounded with probability one. By the Gaussian Isoperimetric Inequality (see, for instance, Pollard (2002), p. 279), this implies that the tail probability of $\bar{G}$ decays exponentially. Therefore, with $\eta > 0$, $T^{k/2} E_\delta \mathbf{1}[T^{1/2} \sup_{t \leq T} ||\delta_t|| > T^\eta] \to 0$. Since ϵ is arbitrary, this implies $\rho_1 \xrightarrow{p} 0$.

For ρ_2, note that for any fixed n, by Condition 1 (ID), there exists $T^*(n)$ such that for all $T > T^*(n)$,

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

For any T, let n_T be the largest n such that simultaneously, $T > \sup_{n' \leq n} T^*(n')$, $T^{1/2} K(n^{-1}) > 1$ and $T^{-1/4} n < 1$. Note that $n_T \to \infty$, since for any fixed n, $T^*(n+1)$ and $n+1$ are finite and $K((n+1)^{-1}) > 0$. By construction,

$$P(\sup_{||\theta - \theta_0|| \geq n_T^{-1}} T^{-1} \sum \sup_{||v|| < T^{-1/2+\eta}, \theta + v \in \Theta} (l_t(\theta + v) - l_t(\theta_0)) < -K(n_T^{-1})) \geq 1 - n_T^{-1}. \tag{23}$$

Now set $a_T = T^{1/2} n_T^{-1} = o(T^{1/2})$. Note that

$$\begin{aligned} \mathcal{S}_T(1 - \mathcal{A}_T) LR_T(u, \delta) &= \mathcal{S}_T(1 - \mathcal{A}_T) \exp[\sum (l_t(\theta_0 + T^{-1/2}u + \delta_t) - l_t(\theta_0))] \\ &\leq (1 - \mathcal{A}_T) \exp[\sum \sup_{||v|| < T^{-1/2+\eta}} (l_t(\theta_0 + T^{-1/2}u + v) - l_t(\theta_0))] \\ &\leq \exp[\sup_{||\theta - \theta_0|| \geq n_T^{-1}} \sum \sup_{||v|| < T^{-1/2+\eta}} (l_t(\theta + v) - l_t(\theta_0))]. \end{aligned}$$

Hence, with probability of at least $1 - n_T^{-1} \to 1$,

$$\begin{aligned} \rho_2 &\leq \int w(\theta_0 + T^{-1/2}u) du \cdot \exp\left[\sup_{||\theta - \theta_0|| \geq n_T^{-1}} \sum \sup_{||v|| < T^{-1/2+\eta}} (l_t(\theta + v) - l_t(\theta_0))\right] \\ &\leq T^{k/2} \exp\left[-TK(n_T^{-1})\right] \leq T^{k/2} \exp\left[-T^{1/2}\right] \to 0 \end{aligned}$$

where the last inequality holds since $T^{1/2} K(n_T^{-1}) > 1$ by construction of n_T.

(ii) Similarly to the reasoning concerning ρ_1 in the proof of part (i), for any $\epsilon > 0$, by Markov's inequality

$$\begin{aligned} P(E_\delta(1-\mathcal{S}_T)LR_T(0,\delta-e\bar{\delta}) &> \epsilon) \leq \epsilon^{-1}EE_\delta(1-\mathcal{S}_T)LR_T(0,\delta-e\bar{\delta}) \\ &= E_\delta(1-\mathcal{S}_T)\int f_T(\theta_0,\delta-e\bar{\delta})d\mu_T \\ &\leq E_\delta(1-\mathcal{S}_T) \to 0. \end{aligned}$$

■

Lemma 4 *Let $D_h^1 = \text{diag}(h_1^1, \cdots, h_T^1)$, $D_h^2 = \text{diag}(h_1^2, \cdots, h_T^2)$ where the $k \times k$ matrices h_t^1 and h_t^2 satisfy $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]}((h_t^1, h_t^2) - (\Gamma_t, \Gamma_t))|| \xrightarrow{p} 0$. If $\hat{\Xi} - \hat{\Gamma}^{-1} \xrightarrow{p} 0$, then*

(i) $\sup_{i,j\leq T} T^{-1}||[F'D_h^1 e\hat{\Xi}e'D_h^2 F - F'D_\Gamma e\hat{\Gamma}^{-1}eD_\Gamma F]_{i,j}|| \xrightarrow{p} 0$

(ii) $\sup_{i,j\leq T} ||[F'(D_h^1 - D_h^2)F]_{i,j}|| \xrightarrow{p} 0$.

Proof. (i) We compute

$$\begin{aligned} & ||[F'D_h^1 e\hat{\Xi}e'D_h^2 F - F'D_\Gamma e\hat{\Gamma}^{-1}e'D_\Gamma F]_{i,j}|| \\ \leq\; & ||[F'D_h^1 e\hat{\Xi}e'D_h^2 F - F'D_\Gamma e\hat{\Xi}e'D_h^2 F]_{i,j}|| \\ & \qquad +||[F'D_\Gamma e\hat{\Gamma}^{-1}e'D_\Gamma F - F'D_\Gamma e\hat{\Gamma}^{-1}e'(I_T\otimes\hat{\Gamma}\hat{\Xi})D_h^2 F]_{i,j}|| \\ =\; & ||[F'(D_h^1 - D_\Gamma)e\hat{\Xi}e'D_h^2 F]_{i,j}|| + ||[F'D_\Gamma e\hat{\Gamma}^{-1}e'(D_\Gamma - (I_T\otimes\hat{\Gamma}\hat{\Xi})D_h^2)F]_{i,j}|| \end{aligned}$$

and

$$\sup_{i,j\leq T} T^{-1}||[F'(D_h^1 - D_\Gamma)e\hat{\Xi}e'D_h^2 F]_{i,j}|| = \sup_{i,j\leq T} ||(T^{-1}\sum_{s=i}^{T}(h_s^1 - \Gamma_s))\hat{\Xi}T^{-1}\sum_{s=j}^{T} h_s^2|| \xrightarrow{p} 0$$

and

$$\begin{aligned} &\sup_{i,j\leq T} T^{-1}||[F'D_\Gamma e\hat{\Gamma}^{-1}e'(D_\Gamma - (I_T\otimes\hat{\Gamma}\hat{\Xi})D_h^2)F]_{i,j}|| \\ &\qquad\qquad = \sup_{i,j\leq T} ||(T^{-1}\sum_{s=j}^{T}\Gamma_s)'\hat{\Gamma}^{-1}T^{-1}\sum_{s=i}^{T}(\Gamma_s - \hat{\Gamma}\hat{\Xi}h_s^2)|| \xrightarrow{p} 0. \end{aligned}$$

(ii) We compute

$$||[F'(D_h^1 - D_h^2)F]_{i,j}|| \leq ||[F'(D_h^1 - D_\Gamma)F]_{i,j}|| + ||[F'(D_h^2 - D_\Gamma)F]_{i,j}||$$

and

$$\begin{aligned} \sup_{i,j\leq T} ||[F'(D_h^1 - D_\Gamma)F]_{i,j}|| &= \sup_{i,j\leq T} T^{-1}||\sum_{t=i\vee j}^{T}(h_t^1 - \Gamma_t)|| \xrightarrow{p} 0 \\ \sup_{i,j\leq T} ||[F'(D_h^2 - D_\Gamma)F]_{i,j}|| &= \sup_{i,j\leq T} T^{-1}||\sum_{t=i\vee j}^{T}(h_t^2 - \Gamma_t)|| \xrightarrow{p} 0. \end{aligned}$$

■

Lemma 5 *Let $\Sigma_\Xi(u)$ be a $Tk \times Tk$ matrix consisting of $k \times k$ blocks $\Xi_{i,j}(u)$, $i, j = 1, \cdots, T$, possibly dependent on u and define $c_T^U = \sup_{i,j \leq T, u \in \mathbb{R}^k} ||\Xi_{i,j}(u)||$. Under Condition 2, there exists a constant c_G independent of u and T such that*

(i) $|\operatorname{tr}(F^{-1}\Sigma_\delta F'^{-1})\Sigma_\Xi(u)| \leq c_T^U c_G$

(ii) $|\operatorname{tr}(F^{-1}\Sigma_\delta F'^{-1})\Sigma_\Xi(u)(F^{-1}\Sigma_\delta F'^{-1})\Sigma_\Xi(u)| \leq (c_T^U)^2 c_G^2$.

Proof. Note that for $1 < i \leq j$, the i,jth $k \times k$ block of $F^{-1}\Sigma_\delta F'^{-1}$ is given by

$$\kappa_G(i/T, j/T) - \kappa_G((i-1)/T, j/T) + \kappa_G((i-1)/T, (j-1)/T) - \kappa_G(i/T, (j-1)/T)$$

by $\kappa_G(1/T, j/T) - \kappa_G(1/T, (j-1)/T)$ for $i = 1 < j$, and by $\kappa_G(1/T, 1/T)$ for $i = j = 1$.

If $i = j$ and $((i-1)/T, i/T] \cap \tau = \varnothing$, due to the symmetry of κ_G and by the Fundamental Theorem of Calculus

$$\begin{aligned} ||\kappa_G(i/T, i/T) - \kappa_G((i-1)/T, i/T)|| &\leq T^{-1} \sup_{r,s \in [0,1] \backslash \tau} ||\nabla_1^- \kappa_G(r,s)|| \\ ||\kappa_G(i/T, (i-1)/T) - \kappa_G((i-1)/T, (i-1)/T)|| &\leq T^{-1} \sup_{r,s \in [0,1] \backslash \tau} ||\nabla_1^+ \kappa_G(r,s)|| \end{aligned}$$

where $\nabla_1^- \kappa_G(r,s)$ and $\nabla_1^+ \kappa_G(r,s)$ are the left and right partial derivatives of κ_G with respect to the first argument, so that in this case, the i,ith block has a norm that is bounded by $T^{-1}c_D = T^{-1}(\sup_{r,s \in [0,1] \backslash \tau} ||\nabla_1^- \kappa_G(r,s)|| + \sup_{r,s \in [0,1] \backslash \tau} ||\nabla_1^+ \kappa_G(r,s)||)$.

If $((j-1)/T, j/T] \cap \tau \neq \varnothing$ and $((i-1)/T, i/T] \cap \tau \neq \varnothing$, then

$$\begin{aligned} &||\kappa_G(i/T, j/T) - \kappa_G((i-1)/T, j/T) + \kappa_G((i-1)/T, (j-1)/T) - \kappa_G(i/T, (j-1)/T)|| \\ &\qquad \leq 4 \sup_{r,s \in [0,1]} ||\kappa_G(r,s)|| = c_J \end{aligned}$$

which is also a valid bound for $||\kappa_G(1/T, 1/T)||$.

If $1 < i < j$ and $((j-1)/T, j/T] \cap \tau = ((i-1)/T, i/T] \cap \tau = \varnothing$, then by the Fundamental Theorem of Calculus

$$\begin{aligned} &||\kappa_G(i/T, j/T) - \kappa_G((i-1)/T, j/T) + \kappa_G((i-1)/T, (j-1)/T) - \kappa_G(i/T, (j-1)/T)|| \\ &\qquad \leq T^{-2} \sup_{r \neq s, r,s \in [0,1] \backslash \tau} ||\frac{\partial^2 \kappa_G(r,s)}{\partial r \partial s}|| = T^{-2} c_O. \end{aligned}$$

Also, if $1 < i < j$ and $((j-1)/T, j/T] \cap \tau \neq \varnothing$, and $((i-1)/T, i/T] \cap \tau = \varnothing$, then by the Fundamental Theorem of Calculus

$$\begin{aligned} ||\kappa_G(i/T, j/T) - \kappa_G((i-1)/T, j/T)|| &\leq T^{-1} \sup_{s \in \tau,\ r \in [0,1] \backslash \tau} ||\frac{\partial \kappa_G(r,s)}{\partial r}|| \\ ||\kappa_G(i/T, (j-1)/T) - \kappa_G((i-1)/T, (j-1)/T)|| &\leq T^{-1} \sup_{s \in \tau,\ r \in [0,1] \backslash \tau} ||\frac{\partial \kappa_G(r,s)}{\partial r}|| \end{aligned}$$

so that the norm of the i,jth block is bounded by $T^{-1}c_C = 2T^{-1}\sup_{s\in\tau,\ r\in[0,1]\backslash\tau}||\frac{\partial\kappa_G(r,s)}{\partial r}||$, which is also a valid bound for $||\kappa_G(1/T,j/T)-\kappa_G(1/T,(j-1)/T)||$.

We can hence decompose

$$TF^{-1}\Sigma_\delta F'^{-1} = \Sigma_D + \Sigma_O + \Sigma_C + \Sigma_J$$

where Σ_D is a block diagonal matrix whose i,ith $k\times k$ block has a norm that is bounded by c_D ("the variance of the increments of the continuous part of δ"), Σ_O is a $Tk\times Tk$ matrix whose i,jth block has a norm that is bounded by $T^{-1}c_O$ ("the covariance of the increments of the continuous part of δ"), $\Sigma_C = \sum_{l=1}^{q}\Sigma_{C,l}$ with $\Sigma_{C,i}$ $Tk\times Tk$ matrices whose only nonzero $k\times k$ blocks are in one (block) row and column and correspond to the jump at time τ_i, and these nonzero blocks have a norm that is bounded by c_C ("the covariance between the jumps and the increments of δ") and Σ_J with q^2 nonzero $k\times k$ blocks whose norm is bounded by $c_J T$ ("the variance of the jumps"), and all these bounds are uniform in i,j and T.

Let A and B be $Tk\times Tk$ matrices with i,jth $k\times k$ block $[A]_{i,j}$ and $[B]_{i,j}$, respectively. Note that the i,jth $k\times k$ block of AB, $[AB]_{i,j}$ satisfies

$$||[AB]_{i,j}|| = ||\sum_{l=1}^{T}A_{i,l}B_{l,j}|| \le \sum_{l=1}^{T}||A_{i,l}||\cdot||B_{l,j}|| \le [\bar{A}\bar{B}]_{i,j}$$

where for any $Tk\times Tk$ matrix C with i,jth $k\times k$ block $[C]_{i,j}$, $\bar{C}$ denotes a $T\times T$ matrix whose i,jth element $[\bar{C}]_{i,j}$ is at least as large as $\mathbf{1}[||[C]_{i,j}||>0]\sup_{s\le T,t\le T}||[C]_{s,t}||$. Also

$$\begin{aligned}
||[ABCB]_{i,j}|| &= ||\sum_{l=1}^{T}[AB]_{i,l}[CB]_{l,j}|| \le \sum_{l=1}^{T}||[AB]_{i,l}||\cdot||[CB]_{l,j}|| \\
&\le \sum_{l=1}^{T}[\bar{A}\bar{B}]_{i,l}[\bar{C}\bar{B}]_{l,j} = [(\bar{A}\bar{B})(\bar{C}\bar{B})]_{i,j}.
\end{aligned}$$

Hence, using $|\operatorname{tr}[AB]_{i,i}| \le k||[AB]_{i,i}||$, we obtain

$$|\operatorname{tr}AB| \le k\operatorname{tr}\bar{A}\bar{B} \quad \text{and} \quad |\operatorname{tr}ABCB| \le k\operatorname{tr}\bar{A}\bar{B}\bar{C}\bar{B}.$$

Note that we can choose $\bar{\Sigma}_O = T^{-1}c_O e_0 e_0'$, $\bar{\Sigma}_D = c_D I_T$, $\bar{\Sigma}_J = Tc_J\iota_\tau\iota_\tau'$, $\bar{\Sigma}_C = c_C(\iota_\tau e_0' + e_0\iota_\tau')$ and $\bar{\Sigma}_\Xi(u) = c_U e_0 e_0'$ where ι_τ is a $T\times 1$ vector with elements $[\iota_\tau]_j = \mathbf{1}[((j-1)/T, j/T]\cap\tau\neq\varnothing]$ and e_0 is a $T\times 1$ vector of ones.

(i) We compute

$$\begin{aligned}
|\operatorname{tr}(F^{-1}\Sigma_\delta F'^{-1})\Sigma_\Xi(u)| &= T^{-1}|\operatorname{tr}(\Sigma_D+\Sigma_O+\Sigma_C+\Sigma_J)\Sigma_\Xi(u)| \\
&\le kT^{-1}\operatorname{tr}(\bar{\Sigma}_D+\bar{\Sigma}_O+\bar{\Sigma}_C+\bar{\Sigma}_J)\bar{\Sigma}_\Xi(u) \\
&= kc_T^U T^{-1}\operatorname{tr}(c_D I_T + T^{-1}c_O e_0 e_0' + Tc_J\iota_\tau\iota_\tau' + c_C(\iota_\tau e_0' + e_0'\iota_\tau))e_0 e_0' \\
&= kc_T^U(c_D + c_O + c_J q^2 + 2c_C q).
\end{aligned}$$

(ii) We compute

$$
\begin{aligned}
& |\operatorname{tr}(F^{-1}\Sigma_\delta F'^{-1})\Sigma_\Xi(u)(F^{-1}\Sigma_\delta F'^{-1})\Sigma_\Xi(u)| \\
= \; & T^{-2}\operatorname{tr}(\Sigma_D+\Sigma_O+\Sigma_C+\Sigma_J)\Sigma_\Xi(u)(\Sigma_D+\Sigma_O+\Sigma_C+\Sigma_J)\Sigma_\Xi(u) \\
\leq \; & T^{-2}k\operatorname{tr}(\bar\Sigma_D+\bar\Sigma_O+\bar\Sigma_C+\bar\Sigma_J)\bar\Sigma_\Xi(u)(\bar\Sigma_D+\bar\Sigma_O+\bar\Sigma_C+\bar\Sigma_J)\bar\Sigma_\Xi(u) \\
= \; & T^{-2}k(c_T^U e_0'[c_D I_T+T^{-1}c_O e_0 e_0'+Tc_J\iota_\tau\iota_\tau'+c_C(\iota_\tau e_0'+e_0'\iota_\tau)]e_0)^2.
\end{aligned}
$$

■

Lemma 6 *Under Conditions 1 and 2:*

(i) There exists a sequence of random variables $\tilde C_T=O_p(1)$ satisfying $\tilde C_T^{-1}=O_p(1)$ such that

$$\sup_{v\in\mathbb{R}^k,T}(E_\delta\exp[-2(\delta-T^{-1/2}ev)'D_{\tilde h}(\delta-T^{-1/2}ev)]-\exp[-\tfrac{1}{2}\tilde C_T||v||^2])\leq 0.$$

(ii) If $\xi=(\xi_1',\cdots,\xi_T')'$ and $D_\zeta(u)=\operatorname{diag}(\zeta_1(u),\cdots,\zeta_T(u))$, where the $k\times 1$ vectors ξ_t and $k\times k$ matrices $\zeta_t(u)$ satisfy $\sup_{t\leq T}||T^{-1/2}\sum_{s=1}^t\xi_s||\xrightarrow{p}0$ and $\sup_{t\leq T,u\in\mathbb{R}^k}||T^{-1}\sum_{s=1}^t\zeta_s(u)||\xrightarrow{p}0$, and if $F'\Sigma_\varepsilon F$ is a $Tk\times Tk$ symmetric matrix consisting of $k\times k$ blocks $[F'\Sigma_\varepsilon F]_{i,j}$, $i,j=1,\cdots,T$ that satisfy $\sup_{i,j\leq T}||[F'\Sigma_\varepsilon F]_{i,j}||\xrightarrow{p}0$, then

$$\underline{\kappa}_T\exp[\underline{\Delta}_T||v||^2]\leq E_\delta\exp[\xi'\delta+T^{-1/2}v'e'D_\zeta(u)\delta-\tfrac{1}{2}\delta'\Sigma_\varepsilon\delta]\leq\overline{\kappa}_T\exp[\bar\Delta_T||v||^2]$$

uniformly in v and T, where the scalar random variables $\underline{\kappa}_T$, $\underline{\Delta}_T$, $\overline{\kappa}_T$ and $\bar\Delta_T$ do not depend on u or v and $\underline{\kappa}_T\xrightarrow{p}1$, $\overline{\kappa}_T\xrightarrow{p}1$, $\underline{\Delta}_T\xrightarrow{p}0$ and $\bar\Delta_T\xrightarrow{p}0$.

(iii) $E_\delta\exp[4\hat s'\delta]=O_p(1)$.

(iv) If $J_T\in\mathcal{D}$ is a nonstochastic sequence converging to $J\in\mathcal{D}$, where $\mathcal{D}$ is the set of cadlag functions on the unit interval, then

$$\sup_T E_\delta\exp[T^{1/2}J_T(1)'(\delta_T-\bar\delta)-T^{1/2}\sum J_T((t-1)/T)'(\delta_t-\delta_{t-1})]<\infty.$$

Proof. (i) A direct calculation yields

$$\frac{E_\delta\exp[-2(\delta-T^{-1/2}ev)'D_{\tilde h}(\delta-T^{-1/2}ev)]}{E_\delta\exp[-2\delta'D_{\tilde h}\delta]}=\exp[-8T^{-1}v'e'(D_{\tilde h}^{-1}+\Sigma_\delta)^{-1}ev].$$

We have $e'(D_{\tilde h}^{-1}+\Sigma_\delta)^{-1}e=e'D_{\tilde h}^{1/2}(I_{Tk}+D_{\tilde h}^{1/2}\Sigma_\delta D_{\tilde h}^{1/2})^{-1}D_{\tilde h}^{1/2}e$, so that

$$
\begin{aligned}
T^{-1}||e'(D_{\tilde h}^{-1}+\Sigma_\delta)^{-1}e|| &\leq T^{-1}||e'D_{\tilde h}e|| \\
T^{-1}||e'(D_{\tilde h}^{-1}+\Sigma_\delta)^{-1}e|| &\geq ||(T^{-1}e'D_{\tilde h}e)^{-1}||^{-1}(1+||D_{\tilde h}^{1/2}\Sigma_\delta D_{\tilde h}^{1/2}||)^{-1}
\end{aligned}
$$

where $T^{-1}e'D_{\tilde h}e\xrightarrow{p}\int\Gamma(s)ds$ and

$$
\begin{aligned}
||D_{\tilde h}^{1/2}\Sigma_\delta D_{\tilde h}^{1/2}|| &\leq \operatorname{tr}D_{\tilde h}\Sigma_\delta \\
&\leq k\sup_{s\in[0,1]}||\kappa_G(s,s)||T^{-1}\operatorname{tr}D_{\tilde h}\xrightarrow{p}k\sup_{s\in[0,1]}||\kappa_G(s,s)||\operatorname{tr}\textstyle\int\Gamma(s)ds
\end{aligned}
$$

and finally $E_\delta \exp[-2\delta' D_{\tilde{h}} \delta] \leq 1$ a.s.

(ii) Let $U = T^{-1/2} F' D_\zeta(u)' ev$. We first show the result for the upper bound. By the Cauchy-Schwarz inequality

$$\begin{aligned} E_\delta \exp[\xi'\delta + \delta' F'^{-1} U - \tfrac{1}{2}\delta'\Sigma_\varepsilon \delta] &\leq (E_\delta \exp[4\delta'\xi])^{1/4}(E_\delta \exp[4\delta' F'^{-1} U])^{1/4}(E_\delta \exp[-\delta'\Sigma_\varepsilon\delta])^{1/2} \\ &= \exp[2\xi'\Sigma_\delta\xi + 2U'F^{-1}\Sigma_\delta F'^{-1}U](E_\delta \exp[-\delta'\Sigma_\varepsilon\delta])^{1/2}. \end{aligned}$$

Now

$$\begin{aligned} U'F^{-1}\Sigma_\delta F'^{-1}U &= T^{-1}v'e'D_\zeta(u)FF^{-1}\Sigma_\delta F'^{-1}F'D_\zeta(u)'ev \\ &\leq ||v||^2 \operatorname{tr} T^{-1}F'D_\zeta(u)ee'D_\zeta(u)'FF^{-1}\Sigma_\delta F'^{-1}. \end{aligned}$$

But the norm of the i,jth $k \times k$ block of $T^{-1}F'D_\zeta(u)ee'D_\zeta(u)'F$ is bounded by $(\sup_{t\leq T, u\in\mathbb{R}^k} ||T^{-1}\sum_{s=t}^T \zeta_s(u)||)^2 \xrightarrow{p} 0$. Hence, by Lemma 5 (i), $\bar{\Delta}_T = \sup_{u\in\mathbb{R}^k} \operatorname{tr}(T^{-1}F'D_\zeta(u)ee'D_\zeta(u)'F)F^{-1}\Sigma_\delta F'^{-1} \xrightarrow{p} 0$, and $U'F^{-1}\Sigma_\delta F'^{-1}U \leq \bar{\Delta}_T ||v||^2$. Similarly, also $\xi'\Sigma_\delta\xi = \operatorname{tr} F'\xi\xi'FF^{-1}\Sigma_\delta F'^{-1} \xrightarrow{p} 0$.

For each T, let A_G be the $(Tk \times \ell)$ matrix such that $F^{-1}\Sigma_\delta F'^{-1}A_G = 0$, and B_A the $Tk \times (Tk - \ell)$ matrix such that $B_A'B_A = I_{Tk-\ell}$ and $B_AB_A' = M_A = I_{kT} - A_G(A_G'A_G)^{-1}A_G'$ (if $F^{-1}\Sigma_\delta F'^{-1}$ is full rank, define $B_A = I_{Tk}$). Then

$$E_\delta \exp[-\delta'\Sigma_\varepsilon\delta] = E_\delta \exp[-\delta' F'^{-1}B_AB_A'F'\Sigma_\varepsilon FB_AB_A'F^{-1}\delta].$$

Note that the covariance matrix of $B_A'F^{-1}\delta$, $B_A'F^{-1}\Sigma_\delta F^{-1\prime}B_A$ is positive definite, and $M_AF^{-1}\Sigma_\delta F^{-1\prime}M_A = F^{-1}\Sigma_\delta F^{-1\prime}$. Let λ_i, $i = 1, \cdots, kT - \ell$ be the eigenvalues of the symmetric matrix

$$\Sigma_S = (B_A'F^{-1}\Sigma_\delta F^{-1\prime}B_A)^{1/2}B_A'F'\Sigma_\varepsilon FB_A(B_A'F^{-1}\Sigma_\delta F^{-1\prime}B_A)^{1/2}.$$

Then, by Lemma 5,

$$\begin{aligned} \sum_{i=1}^{Tk-\ell} \lambda_i &= \operatorname{tr} B_A'F^{-1}\Sigma_\delta F^{-1\prime}B_AB_A'F'\Sigma_\varepsilon FB_A \\ &= \operatorname{tr} F^{-1}\Sigma_\delta F^{-1\prime}F'\Sigma_\varepsilon F \xrightarrow{p} 0 \end{aligned} \tag{24}$$

and also

$$\begin{aligned} \sum_{i=1}^{Tk-\ell} \lambda_i^2 &= \operatorname{tr} B_A'F^{-1}\Sigma_\delta F'^{-1}B_AB_A'F'\Sigma_\varepsilon FB_AB_A'F^{-1}\Sigma_\delta F'^{-1}B_AB_A'F'\Sigma_\varepsilon FB_A \\ &= \operatorname{tr} F^{-1}\Sigma_\delta F'^{-1}F'\Sigma_\varepsilon FF^{-1}\Sigma_\delta F'^{-1}F'\Sigma_\varepsilon F \xrightarrow{p} 0. \end{aligned} \tag{25}$$

Let $\mathcal{L}_T = \mathbf{1}[\sup_{i\leq kT-\ell} |\lambda_i| \leq 1/2]$, and define $\tilde{\Sigma}_\varepsilon = \mathcal{L}_T\Sigma_\varepsilon$, $\tilde{\Sigma}_S = \mathcal{L}_T\Sigma_S$ and $\tilde{\lambda}_i = \mathcal{L}_T\lambda_i$, $i = 1, \cdots, Tk - \ell$. Note that $E(1 - \mathcal{L}_T) \leq P((\sum_{i=1}^{Tk-\ell} \lambda_i^2)^{1/2} > 1/2) \to 0$ by (25), so that it suffices to show $\mathcal{L}_T E_\delta \exp[-\delta'\Sigma_\varepsilon\delta] \xrightarrow{p} 1$. We compute

$$\begin{aligned}
\mathcal{L}_T E_\delta \exp[-\delta'\Sigma_\varepsilon\delta] &\leq E_\delta \exp[-\delta'\tilde{\Sigma}_\varepsilon\delta] \\
&= E_\delta \exp[-\delta' F'^{-1} B_A B_A' F' \tilde{\Sigma}_\varepsilon F B_A B_A' F^{-1}\delta] \\
&= |B_A' F^{-1}\Sigma_\delta F'^{-1} B_A|^{-1/2} |B_A' F' \tilde{\Sigma}_\varepsilon F B_A + (B_A' F^{-1}\Sigma_\delta F'^{-1} B_A)^{-1}|^{-1/2} \\
&= |I_{Tk-\ell} + \tilde{\Sigma}_S|^{-1/2}.
\end{aligned}$$

Since for $x \in [-1/2, 1/2]$, $x - x^2 \leq \ln(1+x) \leq x$, we find

$$\sum_{i=1}^{Tk-\ell} (\tilde{\lambda}_i - \tilde{\lambda}_i^2) \leq \sum_{i=1}^{Tk-\ell} \ln(1+\tilde{\lambda}_i) = \ln|I_{Tk-\ell} + \tilde{\Sigma}_S| \leq \sum_{i=1}^{Tk-\ell} \tilde{\lambda}_i$$

and the result follows from (24) and (25).

For the lower bound, note that by Jensen's inequality,

$$E_\delta \exp[\xi'\delta + T^{-1/2} v' e' D_\zeta(u)\delta - \tfrac{1}{2}\delta'\Sigma_\varepsilon\delta] \geq (E_\delta \exp[-\xi'\delta - T^{-1/2} v' e' D_\zeta(u)\delta + \tfrac{1}{2}\delta'\Sigma_\varepsilon\delta])^{-1}$$

and proceeding as for the upper bound yields the result.

(iii) Note that

$$E_\delta \exp[4 \textstyle\sum s_t(\hat{\theta})'\delta_t] = \exp[8 \operatorname{tr} F'\hat{s}\hat{s}'F(F^{-1}\Sigma_\delta F^{-1\prime})].$$

The i,jth $k\times k$ block of $F'\hat{s}\hat{s}'F$ is given by $(T^{-1/2}\sum_{t=i}^T s_t(\hat{\theta}))(T^{-1/2}\sum_{t=j}^T s_t(\hat{\theta}))'$, whose norm is $O_p(1)$ uniformly in i,j by Lemma 2 (iv). Hence applying Lemma 5 yields $E_\delta \exp[4\sum s_t(\hat{\theta})'\delta_t] = O_p(1)$.

(iv) Let $\bar{J}_T = (J_T(0)', J_T(1/T)', \cdots, J_T((T-1)/T)')'$ and $\hat{J} = eJ_T(1) - \bar{J}_T - T^{-1}F'D_\Gamma e'\hat{\Gamma}^{-1}J_T(1)$. We compute

$$\begin{aligned}
& E_\delta \exp[T^{1/2} J_T(1)'(\delta_T - \bar{\delta}) - T^{1/2} \textstyle\sum J_T((t-1)/T)'(\delta_t - \delta_{t-1})] \\
= \; & E_\delta \exp[(eJ_T(1) - \bar{J}_T - T^{-1}F'D_\Gamma e'\hat{\Gamma}^{-1}J_T(1))'F^{-1}\delta] \\
= \; & \exp[\tfrac{1}{2} \operatorname{tr} \hat{J}\hat{J}'(F^{-1}\Sigma_\delta F^{-1\prime})].
\end{aligned}$$

But the i,jth $k\times k$ block of $\hat{J}\hat{J}'$ is given by $[J_T(1)(I_k - (T^{-1}\sum_{s=i}^T \Gamma_s)\hat{\Gamma}^{-1}) - J_T((i-1)/T)][J_T(1)(I_k - (T^{-1}\sum_{s=j}^T \Gamma_s)\hat{\Gamma}^{-1}) - J_T((j-1)/T)]'$, whose norm is $O(1)$ uniformly in i,j,T by assumption, so that the result follows from Lemma 5. ■

Lemma 7 *Under Conditions 1 and 2:*

(i) $\int E_\delta \left|\widehat{LR}_T(u,\delta) - \mathcal{A}_T\mathcal{S}_T LR_T(u,\delta)\right| du \xrightarrow{p} 0$

(ii) $E_\delta\overline{LR}_T(\delta) - E_\delta\mathcal{S}_T LR_T(0, \delta - e\bar{\delta}) \xrightarrow{p} 0$

(iii) $E_\delta\overline{LR}_T(\delta) - E_\delta \exp[(\delta - e\bar{\delta})'s_0 - \frac{1}{2}\delta D_\Gamma(\delta - e\bar{\delta})] \xrightarrow{p} 0$

(iv) $E_\delta \exp[-\frac{1}{2}\delta' D_{\tilde{h}}\delta + \frac{1}{2}\delta' D_{\tilde{h}} e(e'D_{\tilde{h}}^{-1}e)^{-1}e'D_{\tilde{h}}\delta] - E_\delta \exp[-\frac{1}{2}\delta D_\Gamma(\delta - e\bar{\delta})] \xrightarrow{p} 0$

Proof. Let $\mathcal{U}_T$ be the indicator of the event that $||\hat{u}|| \leq a_T$. By Lemma 2 (iii), $\hat{u} = O_p(1)$, so that $E\mathcal{U}_T \to 1$. Note that if $\mathcal{U}_T\psi_T \xrightarrow{p} 0$ for some sequence of random variables ψ_T, then also $\psi_T \xrightarrow{p} 0$, so that we may multiply the left-hand side of the expressions in (i)-(iv) by $\mathcal{U}_T$. Let T be large enough such that $\Theta_T = \{\theta : ||\theta-\theta_0|| < 2T^{-1/2}a_T + T^{\eta-1/2} + T^{\eta-1/2}(\sup_{\lambda\in[0,1]} ||\Gamma(\lambda)||)/(\inf_{\lambda\in[0,1]} ||\Gamma(\lambda)||)\} \subset \Theta_0$, so that $\theta_0 + \mathcal{A}_T\mathcal{S}_T\mathcal{U}_T(u - \hat{u} + \delta_t - \bar{\delta}) \in \Theta_0$ for all $t \leq T$.

(i) Let $g_v : [0,1] \mapsto \mathbb{R}$ with $g_v(\lambda) = l_t(\theta_0 + \lambda v) - l_t(\theta_0)$. Note that for $\theta_0 + v \in \Theta_T$, g_v is twice continuously differentiable with $g_v'(\lambda) = v's_t(\theta_0 + \lambda v)$ and $g_v''(\lambda) = -v'h_t(\theta_0 + \lambda v)v$, so that by a first order Taylor expansion in the integral remainder form, $l_t(\theta_0 + v) - l_t(\theta_0) = g_v(1) - g_v(0) = g_v'(0) + \int_0^1 \lambda g_v''(1-\lambda)d\lambda = v's_t(\theta_0) - \frac{1}{2}v'(2\int_0^1 \lambda h_t(\theta_0 + (1-\lambda)v)d\lambda)v$, and similarly, $s_t(\theta_0 + v) = s_t(\theta_0) - (\int_0^1 h_t(\theta_0 + \lambda v)d\lambda)v$. Thus, for $||u|| < a_T$, $T^{1/2}\sup_{t\leq T} ||\delta_t|| < T^\eta$ and $||\hat{u}|| < a_T$

$$\begin{aligned} l_t(\theta_0 + T^{-1/2}u + \delta_t) - l_t(\theta_0 + T^{-1/2}u) &= s_t(\theta_0 + T^{-1/2}u)'\delta_t - \tfrac{1}{2}\delta_t'h_{1,t}(u,\delta)\delta_t \\ l_t(\theta_0 + T^{-1/2}u) - l_t(\theta_0) &= T^{-1/2}u's_t(\theta_0) - \tfrac{1}{2}u'h_{2,t}(u)u \\ s_t(\theta_0 + T^{-1/2}u) &= s_t(\theta_0 + T^{-1/2}\hat{u}) - h_{3,t}(u,\hat{u})T^{-1/2}(u - \hat{u}) \\ s_t(\theta_0) &= s_t(\theta_0 + T^{-1/2}\hat{u}) + h_{4,t}(\hat{u})T^{-1/2}\hat{u} \end{aligned} \tag{26}$$

almost surely, where $h_{1,t}(u,\delta) = 2\int_0^1 \lambda h_t(\theta_0 + T^{-1/2}u + (1-\lambda)\delta_t)d\lambda$, $h_{2,t}(u) = 2\int_0^1 \lambda h_t(\theta_0 + (1-\lambda)T^{-1/2}u)d\lambda$, $h_{3,t}(u,\hat{u}) = \int_0^1 h_t(\theta_0 + \lambda T^{-1/2}(\hat{u} - u))d\lambda$ and $h_{4,t}(\hat{u}) = \int_0^1 h_t(\theta_0 + \lambda T^{-1/2}\hat{u})d\lambda$, $t = 1, \cdots, T$. Define $\{h_{1,t}(u,\delta)\}_{t=1}^T = \{h_t(\theta_0)\}_{t=1}^T$ when $||u|| \geq a_T$ or $T^{1/2}\sup_{t\leq T} ||\delta_t|| > T^\eta$, define $\{h_{2,t}(u)\}_{t=1}^T = \{\tilde{h}_t\}_{t=1}^T$ when $||u|| \geq a_T$, define $\{h_{3,t}(u,\hat{u})\}_{t=1}^T = \{\tilde{h}_t\}_{t=1}^T$ when $||u|| \geq a_T$ or $||\hat{u}|| \geq a_T$, and define $\{h_{4,t}(\hat{u})\}_{t=1}^T = \{\tilde{h}_t\}_{t=1}^T$ when $||\hat{u}|| > a_T$. Further, let $\hat{H}_4(\hat{u}) = T^{-1}\sum h_{4,t}(\hat{u})$ and $\hat{H}_2(u) = T^{-1}\sum h_{2,t}(u)$. For notational convenience, we drop the dependence of $h_{1,t}$, $h_{2,t}$, $\hat{H}_3$ and $\hat{H}_4$ on u, $\hat{u}$ and δ. We these definitions, we have

$$\begin{aligned} \sup_{u\in\mathbb{R}^k,\delta\in\mathbb{R}^{Tk}} & \mathcal{A}_T\mathcal{S}_T\mathcal{U}_T|LR_T(u,\delta) \\ & - \exp[\textstyle\sum \hat{s}_t'\delta_t + T^{-1/2}(\hat{u} - u)'\sum h_{3,t}\delta_t - \frac{1}{2}\sum \delta_t'h_{1,t}\delta_t + \hat{u}'\hat{H}_4u - \frac{1}{2}u'\hat{H}_2u]| = 0 \end{aligned}$$

almost surely.

Let

$$\varsigma_T = (\hat{u} - u)'T^{-1/2}\sum(h_{3,t} - \tilde{h}_t)\delta_t - \tfrac{1}{2}\sum \delta_t'(h_{1,t} - \tilde{h}_t)\delta_t - \tfrac{1}{2}u'(\hat{H}_2 - \tilde{H})u + \hat{u}'(\hat{H}_4 - \tilde{H})u.$$

Now $\sup_{u\in\mathbb{R}^k,\delta\in\mathbb{R}^{Tk}} \mathcal{A}_T\mathcal{S}_T\mathcal{U}_T|LR_T(u,\delta) - \widehat{LR}(u,\delta)\exp\varsigma_T| = 0$ a.s., and by the Cauchy-Schwarz inequality and $\mathcal{U}_T \leq 1$ a.s.

$$\begin{aligned} \mathcal{U}_T\int E_\delta\left|\widehat{LR}_T(u,\delta) - \mathcal{A}_T\mathcal{S}_TLR_T(u,\delta)\right|du &= \int E_\delta\widehat{LR}_T(u,\delta)\,|\mathcal{U}_T - \mathcal{A}_T\mathcal{S}_T\mathcal{U}_T\exp\varsigma_T|\,du \\ &\leq \int [(E_\delta\widehat{LR}_T(u,\delta)^2)(E_\delta(1 - \mathcal{A}_T\mathcal{S}_T\exp\varsigma_T)^2)]^{1/2}du. \end{aligned}$$

We have

$$\begin{aligned}\widehat{LR}_T(u,\delta)^2 &= \exp[2\sum \hat{s}_t'\delta_t - \sum \delta_t'\tilde{h}_t\delta_t + 2T^{-1/2}(\hat{u}-u)'\sum \tilde{h}_t\delta_t - u'\tilde{H}u + 2\hat{u}'\tilde{H}u] \\ &= \exp[2\hat{s}'\delta - (\delta - T^{-1/2}e(\hat{u}-u))'D_{\tilde{h}}(\delta - T^{-1/2}e(\hat{u}-u)) + 2\hat{u}'\tilde{H}\hat{u}]\end{aligned}$$

and by another application of the Cauchy-Schwarz inequality

$$E_\delta \widehat{LR}_T(u,\delta)^2 \leq \exp[2\hat{u}'\tilde{H}\hat{u}](E_\delta \exp[4\hat{s}'\delta])^{1/2}(E_\delta \exp[-2(\delta - T^{-1/2}e(\hat{u}-u))'D_{\tilde{h}}(\delta - T^{-1/2}e(\hat{u}-u))])^{1/2}.$$

By Lemmas 2 (iii) and 6 (iii), $\exp[2\hat{u}'\tilde{H}\hat{u}] = O_p(1)$ and $E_\delta \exp[4\hat{s}'\delta] = O_p(1)$. By Lemma 6 (i),

$$E_\delta \exp[-2(\delta - T^{-1/2}e(\hat{u}-u))'D_{\tilde{h}}(\delta - T^{-1/2}e(\hat{u}-u))] \leq \exp[-\tfrac{1}{2}\tilde{C}_T||\hat{u}-u||^2]$$

where $\tilde{C}_T = O_p(1)$ and $\tilde{C}_T^{-1} = O_p(1)$ and does not depend on u.

Therefore,

$$E_\delta \widehat{LR}_T(u,\delta)^2 \leq O_p(1)\exp[-\tfrac{1}{2}\tilde{C}_T||\hat{u}-u||^2] \tag{27}$$

and with $\Phi(u)$ the cdf of $u \sim \mathcal{N}(\hat{u}, \tilde{C}_T^{-1}I_k)$

$$\begin{aligned} & \int [(E_\delta \widehat{LR}_T(u,\delta)^2)(E_\delta(1-\mathcal{A}_T\mathcal{S}_T\exp\varsigma_T)^2)]^{1/2}du \\ \leq\ & O_p(1)\int \exp[-\tfrac{1}{2}||\hat{u}-u||^2\tilde{C}_T](E_\delta(1-\mathcal{A}_T\mathcal{S}_T\exp\varsigma_T)^2)^{1/2}du \\ =\ & O_p(1)(2\pi)^{k/2}\tilde{C}_T^{-k/2}\int (E_\delta(1-\mathcal{A}_T\mathcal{S}_T\exp\varsigma_T)^2)^{1/2}d\Phi(u) \\ \leq\ & O_p(1)(\int E_\delta(1-\mathcal{A}_T\mathcal{S}_T\exp\varsigma_T)^2 d\Phi(u))^{1/2}\end{aligned}$$

where the inequalities use Jensen's inequality.

In order to show $\int E_\delta(1-\mathcal{A}_T\mathcal{S}_T\exp\varsigma_T)^2 d\Phi(u) \xrightarrow{p} 0$, we first compute the expectation with respect to δ. This is complicated by the fact that $h_{1,t}$ depends on δ. To circumvent this problem, we bound ς_T by $\underline{\varsigma}_T \leq \varsigma_T \leq \overline{\varsigma}_T$, where $\underline{\varsigma}_T$ and $\overline{\varsigma}_T$ are defined just as ς_T, but with $h_{1,t}$ replaced by a term that does not depend on δ (or u).

Specifically, for each $t \leq T$, define

$$d_t = 2\sup_{||v||<a_T+T^\eta}\left\|h_t(\theta_0 + T^{-1/2}v) - h_t(\theta_0)\right\|.$$

Note that for any $v \in \mathbb{R}^k$ with $||v|| = 1$,

$$|v'(h_t(\theta_0) - h_{1,t})v| \leq ||h_t(\theta_0) - h_{1,t}|| \leq d_t$$

since for $||u|| < a_T$ and $T^{1/2}\sup_{t\leq T}||\delta_t|| < T^\eta$, $||h_t(\theta_0) - h_{1,t}|| = ||2\int_0^1 \lambda(h_t(\theta_0 + T^{-1/2}u + (1-\lambda)\delta_t) - h_t(\theta_0))d\lambda||$ and $h_{1,t}(u,\delta) = h_t(\theta_0)$ otherwise. Thus, for all $\delta \in \mathbb{R}^{Tk}$,

$$\sum \delta_t'(h_t(\theta_0) - d_tI_k)\delta_t \leq \sum \delta_t'h_{1,t}\delta_t \leq \sum \delta_t'(h_t(\theta_0) + d_tI_k)\delta_t.$$

Now let

$$\begin{aligned}
\overline{\varsigma}_T &= \varsigma_T + \tfrac{1}{2}\sum \delta_t'(h_{1,t} - h_t(\theta_0) + d_t I_k)\delta_t \\
\underline{\varsigma}_T &= \varsigma_T + \tfrac{1}{2}\sum \delta_t'(h_{1,t} - h_t(\theta_0) - d_t I_k)\delta_t
\end{aligned}$$

so that $\underline{\varsigma}_T \leq \varsigma_T \leq \overline{\varsigma}_T$. We obtain

$$\begin{aligned}
0 &\leq E_\delta(1 - \mathcal{A}_T\mathcal{S}_T \exp \varsigma_T)^2 \\
&\leq 1 - 2E_\delta \mathcal{A}_T\mathcal{S}_T \exp \underline{\varsigma}_T + E_\delta \mathcal{A}_T\mathcal{S}_T \exp 2\overline{\varsigma}_T \\
&\leq 1 - 2E_\delta \exp \underline{\varsigma}_T + E_\delta \exp 2\overline{\varsigma}_T + 2E_\delta(1 - \mathcal{A}_T\mathcal{S}_T) \exp \underline{\varsigma}_T.
\end{aligned}$$

We will now show that $\int E_\delta \exp \underline{\varsigma}_T d\Phi(u)$ is bounded below by random variable that converges to one in probability, that $\int E_\delta \exp 2\overline{\varsigma}_T d\Phi(u)$ is bounded above by a random variable that converges to one in probability, and $\int E_\delta(1-\mathcal{S}_T\mathcal{A}_T) \exp \underline{\varsigma}_T d\Phi(u) \xrightarrow{p} 0$, which implies $0 \leq \int E_\delta(1-\mathcal{A}_T\mathcal{S}_T \exp \varsigma_T)^2 d\Phi(u) \xrightarrow{p} 0$.

With $D_{h3} = \mathrm{diag}(h_{3,1}, \cdots, h_{3,T})$, $D_h = \mathrm{diag}(h_1(\theta_0), \cdots, h_T(\theta_0))$ and $D_d = \mathrm{diag}(d_1 I_k, \cdots, d_T I_k)$ we have

$$\begin{aligned}
E_\delta \exp \underline{\varsigma}_T &= \exp[-\tfrac{1}{2}u'(\hat{H}_2 - \tilde{H})u + \hat{u}'(\hat{H}_4 - \tilde{H})u] \\
&\quad \cdot E_\delta \exp[(\hat{u} - u)'T^{-1/2}e'(D_{h3} - D_{\tilde{h}})\delta - \tfrac{1}{2}\delta'(D_h - D_{\tilde{h}} + D_d)\delta]
\end{aligned}$$

and

$$\begin{aligned}
E_\delta \exp 2\overline{\varsigma}_T &= \exp[-u'(\hat{H}_2 - \tilde{H})u + 2\hat{u}'(\hat{H}_4 - \tilde{H})u] \\
&\quad \cdot E_\delta \exp[2(\hat{u} - u)'T^{-1/2}e'(D_{h3} - D_{\tilde{h}})\delta - \delta'(D_h - D_{\tilde{h}} - D_d)\delta].
\end{aligned}$$

Since

$$\begin{aligned}
\sup_{u\in\mathbb{R}^k, t\leq T} T^{-1}||\sum_{s=1}^t (h_{3,s}(u,\hat{u}) - \tilde{h}_s)|| &\leq \sup_{t\leq T, ||u||\leq a_T, ||\hat{u}||<a_T} T^{-1}||\sum_{s=1}^t h_{3,s}(u,\hat{u}) - \tilde{h}_s|| \xrightarrow{p} 0 \\
\sup_{t\leq T} T^{-1}||\sum_{s=1}^t (h_s(\theta_0) + d_t I_k - \tilde{h}_s)|| &\leq \sup_{t\leq T} ||T^{-1}\sum_{s=1}^t (h_s(\theta_0) - \tilde{h}_s)|| + T^{-1}\sum_{t=1}^T d_t \xrightarrow{p} 0
\end{aligned}$$

by (12), Lemma 2 (ii) and Condition 1 (LLLN), and similarly, $\sup_{t\leq T} ||T^{-1}\sum_{s=1}^t (h_s(\theta_0) - d_t I_k - \Gamma_s)|| \xrightarrow{p} 0$, Lemmas 4 (ii) and 6 (ii) are applicable, and we obtain

$$\begin{aligned}
E_\delta \exp \underline{\varsigma}_T &\geq \exp[-\tfrac{1}{2}u'(\hat{H}_2 - \tilde{H})u + \hat{u}'(\hat{H}_4 - \tilde{H})u]\underline{\kappa}_T \exp[\underline{\Delta}_T||u - \hat{u}||^2] \\
E_\delta \exp 2\overline{\varsigma}_T &\leq \exp[-u'(\hat{H}_2 - \tilde{H})u + 2\hat{u}'(\hat{H}_4 - \tilde{H})u]\overline{\kappa}_T \exp[\bar{\Delta}_T||u - \hat{u}||^2]
\end{aligned}$$

uniformly in u, where $\underline{\kappa}_T$, $\overline{\kappa}_T$, $\underline{\Delta}_T$ and $\bar{\Delta}_T$ do not depend on u and $\underline{\kappa}_T \xrightarrow{p} 1$, $\overline{\kappa}_T \xrightarrow{p} 1$, $\underline{\Delta}_T \xrightarrow{p} 0$ and $\bar{\Delta}_T \xrightarrow{p} 0$. Also

$$\sup_{u\in\mathbb{R}^k} ||\hat{H}_2(u) - \tilde{H}|| \leq \sup_{||u||<a_T} T^{-1}||\sum h_{2,t}(u) - \Gamma_t|| \xrightarrow{p} 0$$

by (12) and Lemma 2 (ii), and similarly, $\hat{H}_4 - \hat{\Gamma} \xrightarrow{p} 0$. Thus, $\int E_\delta \exp \underline{\varsigma}_T d\Phi(u) \geq \int \underline{\kappa}'_T \exp[\underline{\Delta}'_T ||u - \hat{u}||^2] d\Phi(u) \xrightarrow{p} 1$ and $\int E_\delta \exp 2\bar{\varsigma}_T d\Phi(u) \leq \int \bar{\kappa}'_T \exp[\bar{\Delta}'_T ||u - \hat{u}||^2] d\Phi(u) \xrightarrow{p} 1$, and we are left to show that $\int E_\delta (1 - \mathcal{S}_T \mathcal{A}_T) \exp \underline{\varsigma}_T d\Phi(u) \xrightarrow{p} 0$. By the Cauchy-Schwarz inequality

$$\left[\int E_\delta (1 - \mathcal{S}_T \mathcal{A}_T) \exp \underline{\varsigma}_T d\Phi(u) \right]^2 \leq \left[\int E_\delta (1 - \mathcal{S}_T \mathcal{A}_T)^2 d\Phi(u) \right] \left[\int E_\delta \exp 2 \underline{\varsigma}_T d\Phi(u) \right].$$

From the same reasoning as above, $\int E_\delta \exp 2\underline{\varsigma}_T d\Phi(u) = O_p(1)$, and

$$\int E_\delta (1 - \mathcal{S}_T \mathcal{A}_T) d\Phi(u) \leq \int E_\delta (1 - \mathcal{S}_T) d\Phi(u) + \int E_\delta (1 - \mathcal{A}_T) d\Phi(u).$$

But $\int E_\delta (1 - \mathcal{S}_T) d\Phi(u) = E_\delta (1 - \mathcal{S}_T) = E_\delta \mathbf{1}[T^{1/2} \sup_{t \leq T} ||\delta_t|| \geq T^\eta] \to 0$, and $\int E_\delta (1 - \mathcal{A}_T) d\Phi(u) \leq \int \mathbf{1}[||u|| \geq a_T] d\Phi(u) \xrightarrow{p} 0$ since $||\hat{u}|| = O_p(1)$, $\tilde{C}_T^{-1/2} = O_p(1)$ and $a_T \to \infty$.

(ii) Similar to the proof of part (i), we have for all $T^{1/2} \sup_{t \leq T} ||\delta_t|| < T^\eta$ and $||\hat{u}|| < a_T$

$$l_t(\theta_0 + \delta_t - e\bar{\delta}) - l_t(\theta_0) = s_t(\theta_0)'(\delta_t - \bar{\delta}) - \tfrac{1}{2}(\delta_t - \bar{\delta})' h_{5,t}(\delta)(\delta_t - \bar{\delta})$$

almost surely for $t = 1, \cdots, T$, where $h_{5,t}(\delta) = 2\int_0^1 \lambda h_t(\theta_0 + (1-\lambda)(\delta_t - \bar{\delta})) d\lambda$ for $T^{1/2} \sup_{t \leq T} ||\delta_t|| < T^\eta$ and $h_{5,t}(\delta) = \tilde{h}_t$. Thus, by (26)

$$\mathcal{S}_T \mathcal{U}_T LR_T(0, \delta - e\bar{\delta}) = \mathcal{S}_T \mathcal{U}_T \exp[\sum s_t(\hat{\theta})' \delta_t + T^{-1/2} \hat{u}' \sum h_{4,t}(\hat{u})(\delta_t - \bar{\delta}) - \tfrac{1}{2} \sum (\delta_t - \bar{\delta})' h_{5,t}(\delta)(\delta_t - \bar{\delta})]$$

almost surely. Define $\hat{H}_5(\delta) = T^{-1} \sum h_{5,t}(\delta)$, and we again omit the dependence of $h_{4,t}(\hat{u})$, $h_{5,t}(\delta)$, $\hat{H}_4(\hat{u})$ and $\hat{H}_5(\delta)$ on $\hat{u}$ and δ for notational convenience.

Let

$$\begin{aligned} \varsigma_T^* &= T^{-1/2} \hat{u}' \sum h_{4,t}(\delta_t - \bar{\delta}) - \tfrac{1}{2} \sum (\delta_t - \bar{\delta})' h_{5,t}(\delta_t - \bar{\delta}) \\ &\quad + \tfrac{1}{2} \sum \delta_t' \tilde{h}_t \delta_t - \tfrac{1}{2} (T^{-1/2} \sum \delta_t' \tilde{h}_t) \tilde{H}^{-1} T^{-1/2} \sum \tilde{h}_t \delta_t. \end{aligned}$$

Now $\sup_{\delta \in \mathbb{R}^{Tk}} \mathcal{S}_T \mathcal{U}_T |LR_T(0, \delta - e\bar{\delta}) - \overline{LR}_T(\delta) \exp \varsigma_T^*| = 0$ a.s., and by the Cauchy-Schwarz inequality and $\mathcal{U}_T \leq 1$ a.s.,

$$\begin{aligned} \mathcal{U}_T E_\delta \left| \overline{LR}_T(\delta) - \mathcal{S}_T LR_T(0, \delta - e\bar{\delta}) \right| &= E_\delta \overline{LR}_T(\delta) \left| \mathcal{U}_T - \mathcal{S}_T \mathcal{U}_T \exp \varsigma_T^* \right| \\ &\leq \left[E_\delta \overline{LR}_T(\delta)^2 \right]^{1/2} \left[E_\delta (1 - \mathcal{S}_T \exp \varsigma_T^*)^2 \right]^{1/2}. \end{aligned}$$

By a direct calculation

$$E_\delta \overline{LR}_T(\delta)^2 = (2\pi)^{-k/2} |2\tilde{H}|^{1/2} \exp[-\hat{u}' \tilde{H} \hat{u}] E_\delta \int \widehat{LR}_T(u, \delta)^2 du$$

which is $O_p(1)$ by (27). Furthermore,

$$\begin{aligned} 0 &\leq E_\delta (1 - \mathcal{S}_T \exp \varsigma_T^*)^2 \\ &= 1 - 2E_\delta \mathcal{S}_T \exp \varsigma_T^* + E_\delta \mathcal{S}_T \exp 2\varsigma_T^* \\ &\leq 1 - 2E_\delta \exp \underline{\varsigma}_T^* + E_\delta \exp 2\bar{\varsigma}_T^* - 2E_\delta (1 - \mathcal{S}_T) \exp \underline{\varsigma}_T^* \end{aligned}$$

where

$$\begin{aligned}
\overline{\varsigma}_T^* &= \varsigma_T^* - \tfrac{1}{2}\sum(\delta_t - \bar{\delta})'(h_t(\theta_0) - 2d_t I_k - h_{5,t})(\delta_t - \bar{\delta}) \\
\underline{\varsigma}_T^* &= \varsigma_T^* - \tfrac{1}{2}\sum(\delta_t - \bar{\delta})'(h_t(\theta_0) + 2d_t I_k - h_{5,t})(\delta_t - \bar{\delta})
\end{aligned}$$

since $\sup_{t\le T}||\delta_t - \bar{\delta}|| \le 2\sup_{t\le T}||\delta_t||$. Let $h_{6,t} = \hat{H}_4\hat{\Gamma}^{-1}\Gamma_t$, and note that

$$\sum h_{4,t}\bar{\delta} = \sum h_{6,t}\delta_t \tag{28}$$

and define $D_{h6} = \mathrm{diag}(h_{6,1}, \cdots, h_{6,T})$ and $D_{hd} = \mathrm{diag}(h_1(\theta_0) - 2d_1 I_k, \cdots, h_T(\theta_0) - 2d_T I_k)$, so that

$$\begin{aligned}
\overline{\varsigma}_T^* &= T^{-1/2}\hat{u}'\sum h_{4,t}(\delta_t - \bar{\delta}) - \tfrac{1}{2}\sum(\delta_t - \bar{\delta})'(h_t(\theta_0) - 2d_t I_k)(\delta_t - \bar{\delta}) \\
&\quad + \tfrac{1}{2}\sum \delta_t'\tilde{h}_t\delta_t - \tfrac{1}{2}(T^{-1/2}\sum \delta_t'\tilde{h}_t)\tilde{H}^{-1}T^{-1/2}\sum \tilde{h}_t\delta_t \\
&= T^{-1/2}\hat{u}'e'(D_{h4} - D_{h6})\delta + \tfrac{1}{2}\delta'[T^{-1}D_{hd}e\hat{\Gamma}^{-1}e'D_\Gamma \qquad (29) \\
&\quad + T^{-1}D_\Gamma e\hat{\Gamma}^{-1}e'D_{hd} - T^{-2}D_\Gamma e\hat{\Gamma}^{-1}e'D_{hd}e\hat{\Gamma}^{-1}e'D_\Gamma - T^{-1}D_{\tilde{h}}e\tilde{H}^{-1}e'D_{\tilde{h}} - D_{hd} + D_{\tilde{h}}]\delta.
\end{aligned}$$

By Lemma 2 (ii), $\sup_{u\in\mathbb{R}^k, t\le T} T^{-1}||\sum_{s=t}^T(h_{4,t} - \hat{H}_4\hat{\Gamma}^{-1}\Gamma_t)|| \xrightarrow{p} 0$, and after adding and subtracting $T^{-1}D_\Gamma e\hat{\Gamma}^{-1}e'D_\Gamma$ twice to the quadratic form in δ in (29) we can appeal to Lemma 4 (i) and (ii) to conclude by Lemma 6 (ii) with $v = \hat{u}$ that $E_\delta \exp 2\overline{\varsigma}_T^* \xrightarrow{p} 1$. By very similar arguments, also $E_\delta \exp \underline{\varsigma}_T^* \xrightarrow{p} 1$. Finally,

$$0 \le (E_\delta(1 - \mathcal{S}_T)\exp\underline{\varsigma}_T^*)^2 \le (E_\delta(1 - \mathcal{S}_T))(E_\delta \exp 2\underline{\varsigma}_T^*)$$

and since $E_\delta \exp 2\underline{\varsigma}_T^* = O_p(1)$, the result follows from $0 \le E_\delta(1 - \mathcal{S}_T) = E_\delta\mathbf{1}[T^{1/2}\sup_{t\le T}||\delta_t|| \ge T^\eta] \to 0$.

(iii) We have for $||\hat{u}|| < a_T$

$$\begin{aligned}
\sum s_t(\theta_0)'(\delta_t - \bar{\delta}) &= \sum s_t(\hat{\theta})'(\delta_t - \bar{\delta}) + T^{-1/2}\hat{u}'\sum h_{4,t}(\hat{u})(\delta_t - \bar{\delta}) \\
&= \sum s_t(\hat{\theta})'\delta_t + T^{-1/2}\hat{u}'\sum(h_{4,t}(\hat{u}) - h_{6,t})\delta_t
\end{aligned}$$

where the second inequality uses $\sum s_t(\hat{\theta}) = 0$ for $||\hat{u}|| < a_T$ and (26). Define

$$\begin{aligned}
\varsigma_T^o &= T^{-1/2}\hat{u}'\sum(h_{4,t} - h_{6,t})\delta_t + \tfrac{1}{2}\sum \delta_t'(\tilde{h}_t - \Gamma_t)\delta_t \qquad (30) \\
&\quad - \tfrac{1}{2}(T^{-1/2}\sum \delta_t'\tilde{h}_t)\tilde{H}^{-1}T^{-1/2}\sum \tilde{h}_t\delta_t + \tfrac{1}{2}(T^{-1/2}\sum \delta_t'\Gamma_t)\hat{\Gamma}^{-1}T^{-1/2}\sum \Gamma_t\delta_t
\end{aligned}$$

so that $\mathcal{U}_T E_\delta \exp[(\delta - e\bar{\delta})'s_0 - \frac{1}{2}\delta D_\Gamma(\delta - e\bar{\delta})] = \mathcal{U}_T E_\delta \overline{LR}_T(\delta)\exp\varsigma_T^o$ a.s. By the Cauchy-Schwarz inequality and $\mathcal{U}_T \le 1$ a.s.

$$(\mathcal{U}_T E_\delta \overline{LR}_T(\delta) - \mathcal{U}_T E_\delta \exp[(\delta - e\bar{\delta})'s_0 - \tfrac{1}{2}\delta D_\Gamma(\delta - e\bar{\delta})])^2 \le (E_\delta \overline{LR}_T^2)E_\delta(1 - 2\exp\varsigma_T^o + \exp 2\varsigma_T^o).$$

But as shown above, $E_\delta \overline{LR}_T^2 = O_p(1)$, and an application of Lemmas 2 (ii), 4 (i) and (ii) and 6 (ii) yields $E_\delta \exp\varsigma_T^o \xrightarrow{p} 1$ and $E_\delta \exp 2\varsigma_T^o \xrightarrow{p} 1$.

(iv) Follows from the same reasoning as in part (iii) with $h_{4,t} - h_{6,t}$ in (30) replaced by 0. ■

In the following lemma, we write $\int G^{*\prime}\Gamma^{1/2}dW$ for $\int_0^1 G^*(s)\Gamma(s)^{1/2}dW(s)$, $\int G^{*\prime}\Gamma G^*$ for $\int_0^1 G^*(s)'\Gamma(s)G^*(s)ds$ and so forth.

Lemma 8 *Under Conditions 1 and 2,*

$$E_\delta \overline{LR}_T(\delta) \Rightarrow E_G \exp[\int G^{*\prime}\Gamma^{1/2}dW - \tfrac{1}{2}\int G^{*\prime}\Gamma G^*]$$

where $G^*(s) = G(s) - (\int \Gamma(\lambda)d\lambda)^{-1}\int \Gamma(\lambda)G(\lambda)d\lambda$, *and the limiting distribution is absolutely continuous.*

Proof. By Lemma 7 (iii),

$$E_\delta \overline{LR}_T(\delta) - E_\delta \exp[(\delta - e\bar{\delta})'s_0 - \tfrac{1}{2}\delta D_\Gamma(\delta - e\bar{\delta})] \xrightarrow{p} 0.$$

Note that, by the summation by parts formula,

$$\sum s_t(\theta_0)'\delta_t = \delta_T' \sum_{t=1}^T s_t(\theta_0) - \sum_{t=1}^T (\sum_{s=1}^{t-1} s_s(\theta_0))'(\delta_t - \delta_{t-1}).$$

Now by Lemma 2 (i), $T^{-1/2}\sum_{t=1}^{[\cdot T]} s_t(\theta_0) \Rightarrow \int_0^\cdot \Gamma(\lambda)^{1/2}dW(\lambda)$, where the convergence is on the space $\mathcal{D}_{[0,1]}$ of cadlag functions on the unit interval in the Skorohod metric. By the Skorohod representation Theorem (see, for instance, Davidson (1994), Theorem 26.25), there exists a sequence of stochastic processes $S_T \in \mathcal{D}_{[0,1]}$ defined on some probability space $(\tilde{\mathcal{F}}, \tilde{\mathfrak{F}}, \tilde{P})$ and event $\tilde{A} \in \tilde{\mathfrak{F}}$ with $\tilde{P}(\tilde{A}) = 1$, such that S_T has the same distribution as $T^{-1/2}\sum_{t=1}^{[\cdot T]} s_t(\theta_0)$, S has the same distribution as $\int_0^\cdot \Gamma(\lambda)^{1/2}dW(\lambda)$ (and is continuous with $S(0) = 0$) and $S_T(\cdot, \tilde{\omega}) \to S(\cdot, \tilde{\omega})$ for all $\tilde{\omega} \in \tilde{A}$. Denote by $(\tilde{\mathcal{F}}_p, \tilde{\mathfrak{F}}_p, \tilde{P}_p)$ the probability space obtained as the product space of $(\tilde{\mathcal{F}}, \tilde{\mathfrak{F}}, \tilde{P})$ and $(\mathcal{F}_G, \mathfrak{F}_G, P_G)$, where G of Condition 2 is a stochastic process defined on $(\mathcal{F}_G, \mathfrak{F}_G, P_G)$ (so that E_δ denotes integration with respect to a measure induced by P_G). By this construction,

$$\begin{aligned}\widetilde{LR}_T(\delta, S_T) \quad &= \quad \exp[T^{1/2}S_T(1)'\delta_T - T^{1/2}\sum S_T((t-1)/T)'(\delta_t - \delta_{t-1}) - T^{1/2}S_T(1)'\bar{\delta} \\ &\qquad -\tfrac{1}{2}\sum \delta_t'\Gamma_t\delta_t + \tfrac{1}{2}(T^{-1/2}\sum \Gamma_t\delta_t)'\hat{\Gamma}^{-1}(T^{-1/2}\sum \Gamma_t\delta_t)]\end{aligned}$$

is a random variable defined on $(\tilde{\mathcal{F}}_p, \tilde{\mathfrak{F}}_p, \tilde{P}_p)$, and $E_\delta\widetilde{LR}_T(\delta)$ defined on $(\tilde{\mathcal{F}}, \tilde{\mathfrak{F}}, \tilde{P})$ and $E_\delta \exp[(\delta - e\bar{\delta})'s_0 - \frac{1}{2}\delta D_\Gamma(\delta - e\bar{\delta})]$ defined on $(\mathcal{F}, \mathfrak{F}, P)$ have the same distribution for all T (since they are functions of S_T and $T^{-1/2}\sum_{t=1}^{[\cdot T]} s_t(\theta_0)$, respectively). It therefore suffices to find the limiting distribution of $E_\delta\widetilde{LR}_T(\delta)$.

With $\bar{S}_T(\tilde{\omega}) = (S_T(0, \tilde{\omega})', S_T(1/T, \tilde{\omega})', \cdots, S_T((T-1)/T, \tilde{\omega})')'$ and $\bar{S}$ defined analogously, $T^{1/2}\sum S_T((t-1)/T, \tilde{\omega})'(\delta_t - \delta_{t-1}) = \bar{S}_T(\tilde{\omega})'F^{-1}\delta$, so that for any $\tilde{\omega} \in \tilde{A}$,

$$E_\delta[(\bar{S}_T(\tilde{\omega}) - \bar{S}(\tilde{\omega}))'F^{-1}\delta\delta'F^{-1\prime}(\bar{S}_T(\tilde{\omega}) - \bar{S}(\tilde{\omega}))] = \operatorname{tr} F^{-1}\Sigma_\delta F^{-1\prime}(\bar{S}_T(\tilde{\omega}) - \bar{S}(\tilde{\omega}))(\bar{S}_T(\tilde{\omega}) - \bar{S}(\tilde{\omega}))'.$$

But the i,jth $k\times k$ block of $(\bar{S}_T(\tilde{\omega})-\bar{S}(\tilde{\omega}))(\bar{S}_T(\tilde{\omega})-\bar{S}(\tilde{\omega}))'$ is equal to

$$(\bar{S}_T((i-1)/T,\tilde{\omega})-\bar{S}((i-1)/T,\tilde{\omega}))(\bar{S}_T((j-1)/T,\tilde{\omega})-\bar{S}((j-1)/T,\tilde{\omega}))'$$

whose norm converges to zero uniformly in i and j. Therefore, by Lemma 5, $(\bar{S}_T(\tilde{\omega})-\bar{S}(\tilde{\omega}))'F^{-1}\delta \xrightarrow{p} 0$ in P_G, and hence

$$\begin{aligned}&\exp[T^{1/2}S_T(1,\tilde{\omega})'(\delta_T-\bar{\delta})-\bar{S}_T(\tilde{\omega})'F^{-1}\delta-\tfrac{1}{2}\delta' D_\Gamma(\delta-e\bar{\delta})]\\ &\qquad -\exp[T^{1/2}S_T(1,\tilde{\omega})'(\delta_T-\bar{\delta})-\bar{S}(\tilde{\omega})'F^{-1}\delta-\tfrac{1}{2}\delta' D_\Gamma(\delta-e\bar{\delta})]\xrightarrow{p}0\end{aligned}$$

in P_G.

By Theorem 21, p. 64, of Protter (2005), and the CMT,

$$\begin{aligned}&\exp[T^{1/2}S_T(1,\tilde{\omega})'(\delta_T-\bar{\delta})-\bar{S}(\tilde{\omega})'F^{-1}\delta-\tfrac{1}{2}\delta' D_\Gamma(\delta-e\bar{\delta})]\\ \Rightarrow\ &\exp[S(1,\tilde{\omega})'(G(1)-(\textstyle\int\Gamma)^{-1}\int\Gamma G)-\int S(l,\tilde{\omega})'dG(l)-\tfrac{1}{2}\int G'\Gamma G+(\int\Gamma G)'(\int\Gamma)^{-1}(\int\Gamma G)]\end{aligned}\tag{31}$$

in P_G. Furthermore,

$$E_\delta(\widetilde{LR}_T(\delta,S_T(\cdot,\tilde{\omega})))^2\le E_\delta\exp[2T^{1/2}S_T(1,\tilde{\omega})'(\delta_T-\bar{\delta})-2\bar{S}_T(\tilde{\omega})'F^{-1}\delta]$$

which is uniformly bounded in T by Lemma 6 (iv), so that for all $\tilde{\omega}\in\tilde{A}$, $\widetilde{LR}_T(\delta,S_T(\cdot,\tilde{\omega}))$ is uniformly integrable on $(\mathcal{F}_G,\mathfrak{F}_G,P_G)$. Hence (31) implies that also

$$\begin{aligned}&E_\delta\widetilde{LR}_T(\delta,S_T(\cdot,\tilde{\omega}))\\ \to\ &E_G\exp[S(1,\tilde{\omega})'(G(1)-(\textstyle\int\Gamma)^{-1}\int\Gamma G)-\int S(l,\tilde{\omega})'dG(l)-\tfrac{1}{2}\int G'\Gamma G+(\int\Gamma G)'(\int\Gamma)^{-1}(\int\Gamma G)].\end{aligned}$$

But almost sure convergence implies convergence in distribution, so that in $(\tilde{\mathcal{F}},\tilde{\mathfrak{F}},\tilde{P})$

$$\begin{aligned}&E_\delta\widetilde{LR}_T(\delta,S_T)\\ \Rightarrow\ &E_G\exp[S(1)'(G(1)-(\textstyle\int\Gamma)^{-1}\int\Gamma G)-\int S'dG-\tfrac{1}{2}\int G'\Gamma G+(\int\Gamma G)'(\int\Gamma)^{-1}(\int\Gamma G)]\\ =\ &E_G\exp[\textstyle\int(G-(\int\Gamma)^{-1}\int\Gamma G)'dS-\tfrac{1}{2}\int(G-(\int\Gamma)^{-1}\int\Gamma G)'\Gamma G]\\ \sim\ &E_G\exp[\textstyle\int G^{*\prime}\Gamma^{1/2}dW-\tfrac{1}{2}\int G^{*\prime}\Gamma G^*]\end{aligned}$$

where the equality follows from the integration by parts formula on p. 83 of Protter (2005).

Finally, conditional on $G^*\neq 0$, $\exp[\int G^{*\prime}\Gamma^{1/2}dW-\frac{1}{2}\int G^{*\prime}\Gamma G^*]$ has a nondegenerate lognormal distribution, which is absolutely continuous. With $E_G\mathbf{1}[G^*\neq 0]=1$, the mixture of lognormals $E_G\exp[\int G^{*\prime}\Gamma^{1/2}dW-\frac{1}{2}\int G^{*\prime}\Gamma G^*]$ is therefore absolutely continuous, too. ■

References

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

——— (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.

——— (2003): "End-of-Sample Instability Tests," *Econometrica*, 71, 1661–1694.

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

Bai, J. (1997): "Estimation of a Change Point in Multiple Regressions," *Review of Economics and Statistics*, 79, 551–563.

Bai, J., R. L. Lumsdaine, and J. H. Stock (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432.

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

Bernanke, B. S., and I. Mihov (1998): "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113, 869–902.

Boivin, J. (2003): "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real Time Data," *Working paper, Columbia University.*

Brown, L. D., and M. G. Low (1996): "Asymptotic Equivalence of Nomparametric Regression and White Noise," *Annals of Statistics*, 24, 2384–2398.

Cai, Z. (2007): "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, 136, 163–188.

Carrasco, M., L. Hu, and W. Ploberger (2005): "Optimal Test for Markov Switching," *Working paper, University of Rochester.*

Chernoff, H., and S. Zacks (1964): "Estimating the Current Mean of a Normal Distribution Which is Subject to Changes in Time," *Annals of Mathematical Statistics*, 35, 999–1028.

Chib, S. (1998): "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 75, 221–241.

Cogley, T., and T. J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8, 262–302.

Cooley, T. F., and E. Prescott (1973): "Varying Parameter Regressions: A Theory and some Applications," *Annals of Economic and Social Measurement*, 2, 463–474.

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

Dufour, J.-M., and E. Ghysels (1996): "Recent Developments in the Econometrics of Structural Change : Overview," *Journal of Econometrics*, 70, 1–8.

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

Elliott, G., and U. K. Müller (2004): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Working paper, Princeton University.*

——— (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73, 907–940.

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

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

Ghysels, E. (1998): "On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?," *Journal of Finance*, 53, 549–573.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and its Applications*. Academic Press, New York.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

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

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Harvey, A. C., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Jacquier, E., N. G. Polson, and P. E. Rossi (1995): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–417.

Jong, P. D. (1991): "The Diffuse Kalman Filter," *The Annals of Statistics*, pp. 1073–1083.

Kim, C.-J., and C. R. Nelson (1999): *State-Space Models with Regime Switching.* The MIT Press, Cambridge, MA.

LeCam, L. (1986): *Asymptotic Methods in Statistical Decision Theory.* Springer Verlag, New York.

Lettau, M., and S. C. Ludvigson (2001): "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56, 815–849.

Li, H., and U. K. Müller (2006): "Valid Inference in Partially Unstable GMM Models," *Working paper, Princeton University.*

Linde, J. (2001): "Testing for the Lucas-Critique: A Quantitative Investigation," *American Economic Review*, 91, 986–1005.

McConnell, M. M., and G. Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980's," *American Economic Review*, 90, 1464–1476.

McLeish, D. L. (1974): "Dependent Central Limit Theorems and Invariance Principles," *The Annals of Probability*, 2, 620–628.

Nussbaum, M. (1996): "Asymptotic Equivalence of Density Estimation and Gaussian White Noise," *Annals of Statistics*, 24, 2399–2430.

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

Paye, B. S., and A. Timmermann (2006): "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

Phillips, P. C. B., and W. Ploberger (1996): "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica*, 64, 381–412.

Phillips, P. C. B., and W. Ploberger (2006): "Optimal Estimation under Nonstandard Conditions," *Working paper, Washington University in Saint Louis.*

Ploberger, W. (2004): "A Complete Class of Tests When the Likelihood is Locally Asymptotically Quadratic," *Journal of Econometrics*, 118, 67–94.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability.* Cambridge University Press, Cambridge, UK.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821–852.

Protter, P. E. (2005): *Stochastic Integration and Differential Equations.* Springer Verlag, Berlin, 2nd edition edn.

Robinson, P. M. (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl, pp. 253–264. Springer, Berlin.

——— (1991): "Time-Varying Nonlinear Regression," in *Economic Structural Change. Analysis and Forecasting*, ed. by P. Hackl, and A. H. Westlund, pp. 179–190, Berlin. Springer-Verlag.

Rosenberg, B. (1973): "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2, 399–428.

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

Shephard, N., and M. K. Pitt (1997): "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, 653–667.

Shively, T. S. (1988a): "An Analysis of Tests for Regression Coefficient Stability," *Journal of Econometrics*, 39, 367–386.

——— (1988b): "An Exact Test for a Stochastic Coefficient in a Time Series Regression Model," *Journal of Time Series Analysis*, 9, 81–88.

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

Sims, C. A., and T. Zha (2006): "Where There Regime Switches in Us Monetary Policy?," *Amercian Economic Review*, 96, 54–81.

Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4, pp. 2740–2841. North Holland, New York.

Stock, J. H., and M. W. 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.

——— (2002): "Has the Business Cycle Changed and Why?," in *NBER Macroeconomics Annual 2002*, ed. by M. Gertler, and K. S. Rogoff, pp. 159–218. The MIT Press, Cambridge, MA.

Uhlig, H. (2001): "Baeysian Vector Autoregressions with Stochastic Volatility," *Econometrica*, 65, 59–73.

Vogelsang, T. J. (1998): "Testing for a Mean Shift Without Having to Estimate Serial Correlation Parameters," *Journal of Business and Economic Statistics*, 16, 73–80.

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.

White, H. (2001): *Asymptotic Theory for Econometricians*. Academic Press, San Diego, revised edn.