Optimal Test for Markov Switching

Marine Carrasco Liang Hu Werner Ploberger

University of Rochester

September 2004
Comments welcome

1. Introduction

The aim of the paper is to propose an optimal test for the null hypothesis of parameter constancy $H_0 : \theta_t = \theta_0$ against an alternative where the parameters vary according to an unobservable Markov chain. This testing problem includes testing the parameter stability in a Markov-switching model (Hamilton, 1989) and in a random coefficient model (for example a state space model). The model under the null need not be linear, it may be a GARCH model for instance.

The parameters driving the dynamic of the underlying Markov chain are not identified under the null hypothesis. As a result, the testing problem is non-standard and the likelihood ratio test does not converge to a chi-square distribution. Our test is based on functionals of expressions like

$$\frac{1}{\sqrt{T}}\sum_t h'\left[\left(\frac{\partial^2 l_t}{\partial\theta\partial\theta'}+\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_t}{\partial\theta}\right)'\right)+2\sum_{s<t}\rho^{(t-s)}\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)'\right]h \qquad (1.1)$$

where l_t denotes the conditional log-likelihood for one observation under H_0 and h and ρ are nuisance parameters (h measures the difference between the states, and ρ measures the autocorrelation of the variations of the parameter θ_t) . This test is strongly related to the Information Matrix test introduced by White (1982). It has the advantage of using the estimation of the model under H_0 only. We show that, for fixed values of the nuisance parameters, our test is equivalent to the likelihood ratio (LR) test. The nuisance parameters are integrated out to obtain an admissible test.

There are few papers proposing tests for Markov-switching. Garcia (1998) studies the asymptotic distribution of a sup-type Likelihood ratio test. Hansen (1992) treats the likelihood as a empirical process indexed by all the parameters (those identified and those unidentified under the null). The test relies on taking the supremum of LR over the nuisance parameters. Both papers require estimating the model under the alternatives, which may be cumbersome. None investigates local powers. Some work has been done on testing for independent mixture, Chesher (1984), Lee and Chesher (1986), Davidson and MacKinnon (1991), and recently Cho and White (2003).

It should be emphasized that testing parameter stability against a Markov switching alternative is much more challenging than testing for Structural change or Threshold. They have in common that they involve nuisance parameters that are not identified under the null hypothesis. The latter have been investigated in many papers: Davies (1977, 1987), Andrews (1993), Andrews and Ploberger (1994), Hansen (1996) among others. There is, however, some difference to the classical situation: the "right" local alternatives are of order $T^{-1/4}$. Hence, to study the properties of this test, we need to do expansions of the likelihood at the fourth order.

The outline of the paper is as follows. Section 2 describes the test statistic. Section 3 establishes the admissibility. In Section 4, we describe simulation results. Finally in Section 5, we use this test to investigate the presence of rational bubbles in stock markets. Using US data of stock price and dividend, we find evidence in favor of periodically collapsing bubbles. In Appendix A, we define the tensor notations used to derive the fourth order expansion of the likelihood. These notations are interesting in their own as they could be used in other econometric problems involving higher-order expansions. The proofs are collected in Appendix B.

2. Assumptions and test statistic

The observations are given by $y_1, y_2, ..., y_T$. Let $f_t(.)$ be the conditional density (with respect to a dominating measure) of y_t given $y_{t-1}, ..., y_1$. Let μ_T be the dominating measure for the density of $(y_1, y_2, ..., y_T)$. We assume that each $f_t(.)$ is indexed by a k-dimensional vector of parameters, say θ_t. We are interested in testing the stability of these parameters, namely we test

$$H_0 : \theta_t = \theta_0, \text{ for some unspecified } \theta_0 \tag{2.1}$$

against

$$H_1 : \theta_t = \theta_0 + \eta_t, \tag{2.2}$$

where the switching variable η_t is not observable.

Assumption 1. (i) η_t is a homogeneous Markov chain, which is stationary and geometric ergodic, that is, there exist $0 < \lambda < 1$ and a measurable non-negative function g such that

$$\sup_{|f|\leq 1} \left| E\left[f\left(\eta_{t+m}\right)|\eta_t\right] - E\left[f\left(\eta_t\right)\right]\right| \leq \lambda^m g\left(\eta_t\right).$$

and $Eg(\eta_t) < \infty$. (ii) $E\eta_t = 0$ and $\max_t \|\eta_t\| \leq M < \infty$. (iii) η_t does not depend on $y_{t-1}, ..., y_1$.

Remark 1. The assumption $E\eta_t = 0$ is not restrictive as the model can always be reparametrized to ensure this condition. η_t geometric ergodic is satisfied by e.g. irreducible and aperiodic Markov chain with finite state space. $\max_t \|\eta_t\| \leq M < \infty$ will also be satisfied by any finite state space Markov chain, however it will not be satisfied by an AR(1) process with normal error. This condition could be relaxed to allow for distributions of η_t with thin tails but this extension is beyond the scope of the present paper.

Assumption 2. The autocorrelations of η_t depend on some unknown parameters β, which includes λ. They are nuisance parameters that are not identified under H_0. β belongs to a set B, and λ is bounded away from 1.

Assumption 3. y_t is stationary under H_0 and the following conditions on the conditional log-density of y_t given $y_{t-1}, ..., y_1$ (under H_0), l_t, are satisfied:

$$\begin{aligned}
\sup_{t,\theta\in\mathcal{N}} E\left(\left\|l_t^{(1)}\right\|^{24}\right) &< \infty, \\
\sup_{t,\theta\in\mathcal{N}} E\left(\left\|l_t^{(2)}\right\|^{12}\right) &< \infty, \\
\sup_{t,\theta\in\mathcal{N}} E\left(\left\|l_t^{(3)}\right\|^{8}\right) &< \infty, \\
\sup_{t,\theta\in\mathcal{N}} E\left(\left\|l_t^{(4)}\right\|\right) &< \infty, \\
\sup_{t,\theta\in\mathcal{N}} E\left(\left\|l_t^{(5)}\right\|\right) &< \infty.
\end{aligned}$$

where $\mathcal{N}$ is a neighborhood around θ_0.

Remark 2. We do not impose restrictions on the moments of y_t. For instance y_t could be a stationary IGARCH process. However, we rule out the case where y_t is a random walk. To deal with unit root, we would have to alter the test statistic by proper rescaling and its asymptotic distribution would be different. We leave this extension for future research.

Note that y_t may include exogenous variables.

The test statistic, for a given β, is of the form.

$$TS_T(\beta) = TS_T\left(\beta, \hat{\theta}\right) = \Gamma_T - \frac{1}{2T}\widehat{\varepsilon}(\beta)'\widehat{\varepsilon}(\beta)$$

where

$$\begin{aligned}
\Gamma_T &= \frac{1}{2}\left(\frac{1}{\sqrt{T}}\sum_t\left(l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)}\right) var(\eta_t) + \frac{2}{\sqrt{T}}\sum_{t>s} l_t^{(1)} \otimes l_s^{(1)} cov(\eta_t, \eta_s)\right) \quad (2.3) \\
&\equiv \frac{1}{2\sqrt{T}}\sum_t \mu_{2,t}\left(\beta, \hat{\theta}\right),
\end{aligned}$$

and $\widehat{\varepsilon}(\beta)$ is the residual from the OLS regression of $\frac{1}{2}\mu_{2,t}\left(\beta, \hat{\theta}\right)$ on $l_t^{(1)}\left(\hat{\theta}\right)$, and $\hat{\theta}$ is the maximum likelihood estimator of θ under H_0 (i.e. the ML estimator under the assumption of constant parameters).

As β is unknown and can not be estimated consistently under H_0, we use sup-type tests like in Davies (1987) and exponential-type tests as in Andrews and Ploberger (1994):

$$\text{supTS} = \sup_{\beta \in \bar{B}} TS_T(\beta)$$

or

$$\text{expTS} = \int_{\bar{B}} \exp\left(TS_T(\beta)\right) dJ(\beta)$$

where J is some prior distribution for β with support on $\bar{B}$ a compact subset of B. We will establish admissibility for a class of expTS statistics.

Remark 3. The asymptotic distribution of the tests will not be nuisance parameter free in general. Therefore we have to rely on parametric bootstrap to compute the critical values.

Remark 4. The test statistic TS depends only on the score and derivative of the score under the null and on the estimator of θ under H_0. Therefore it does not require estimating the model under the alternative. This is a great advantage over competing tests like those of Garcia (1998), Hansen (1992) because estimating a Markov switching model is particularly burdensome as one needs to use the EM algorithm (Hamilton, 1989).

Remark 5. The test relies on the second Bartlett identity (Bartlett, 1953a,b). It is related to the Information Matrix test introduced by White (1982). Chesher (1984) shows the Information Matrix test has power against models with random coefficients. He shows that a score test of the hypothesis that parameters have zero variance is close to the Information Matrix test. Davidson and McKinnon (1991) derive information-matrix-type tests for testing random parameters. The main difference with our setting is that they assume that the parameters are independent, whereas we assume that the parameters are serially correlated and we fully exploit this correlation. Our optimality results do not apply to independent mixture models. Recently, Cho and White (2003) have proposed a test for independent mixture.

Remark 6. The form of our test is insensitive to the dynamic of the latent process η_t. It depends only on the form of the autocorrelation of η_t.

Remark 7. We assume throughout the paper that the model under the null is correctly specified. The issue of misspecification is not addressed here.

Remark 8. The main difference with Structural change and threshold testing is that here the local alternatives are of order $T^{-1/4}$. This is due to the fact that the regimes η_t are unknown and one needs to estimate them at each period.

Although the optimality results are proved under the general assumptions 1 to 3, the expression of the test statistic can be simplified under the following extra assumption.

Assumption 4. η_t can be written as chS_t where S_t is a scalar Markov chain with $V(S_t) = 1$, h is a vector specifying the direction of the alternative (for identification h is normalized so that $\|h\| = 1$), and c is a scalar specifying the amplitude of the change. Moreover, $corr(S_t, S_s) = \rho^{|t-s|}$ for some $-1 < \rho < 1$. In such case, $\beta = (c^2, h', \rho)'$.

Assumptions 1 and 4 impose some restrictions on the Markov chain S_t. If S_t has a finite state space, then it will be geometric ergodic provided its transition probability matrix satisfies some restrictions described e.g. in Cox and Miller (1965, page 124). More precisely, if S_t is a two-state Markov chain, which takes the values a and b, and has transition probabilities $p = P(S_t = a|S_{t-1} = a)$ and $q = P(S_t = b|S_{t-1} = b)$, S_t is geometric ergodic if $0 < p < 1$ and $0 < q < 1$. In this example $\rho = p + q - 1$.

S_t can also have a continuous state space as long as it is bounded. Consider an autoregressive model

$$S_t = \rho S_{t-1} + \varepsilon_t$$

where ε_t is iid $U[-1,1]$ and $-1 < \rho < 1$. Then S_t has bounded support $(-1/(1-|\rho|), 1/(1-|\rho|))$ and has mean zero. Moreover it is easy to check that S_t is geometric ergodic using Theorem 3 page 93 of Doukhan (1994). For this choice of S_t, y_t follows a random coefficient model under the alternative.

Under Assumption 4, $\mu_{2,t}(\beta,\theta)$ can be written as

$$\mu_{2,t}(\beta,\theta) = c^2 h' \left[\left(\frac{\partial^2 l_t}{\partial\theta\partial\theta'} + \left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_t}{\partial\theta}\right)' \right) + 2\sum_{s<t} \rho^{(t-s)} \left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)' \right] h, \qquad (2.4)$$

and $\bar{B} = \left\{c^2, h, \rho : c^2 > 0, \|h\| = 1, \underline{\rho} < \rho < \bar{\rho}\right\}$ and $-1 < \underline{\rho} < \bar{\rho} < 1$.

The maximum of $TS_T(\beta)$ with respect to c^2 can be computed analytically. As a result, we get

$$\text{supTS} = \sup_{\left\{h,\rho:\|h\|=1,\underline{\rho}<\rho<\bar{\rho}\right\}} \frac{1}{2}\left(\max\left(0, \frac{\Gamma_T^*}{\sqrt{\widehat{\varepsilon}^{*\prime}\widehat{\varepsilon}^*}}\right)\right)^2$$

where Γ_T^* is $\Gamma_T(\beta)/c^2$ and $\widehat{\varepsilon}^* = \widehat{\varepsilon}(\beta)/(\sqrt{T}c^2)$ so that Γ_T^* and $\widehat{\varepsilon}^*$ do not depend on c^2.

3. Local alternatives and asymptotic optimality

First of all let us discuss some general principles for the construction of admissible tests. A test is admissible if there is no other test that has uniformly higher (or equal) power. Suppose now we have a general testing problem of testing a null H_0 against an alternative H_1 and let μ_0 and μ_1 be two measures concentrated on H_0 and H_1, respectively. Furthermore assume that the probability measures for our models are given by densities f_ξ, (with respect to a common dominating measure), where the parameter $\xi \in H_0 \cup H_1$. Then a test rejecting when

$$\frac{\int f_\xi d\mu_1}{\int f_\xi d\mu_0} > K \qquad (3.1)$$

are admissible (This is an easy generalization of the Neyman-Pearson lemma: For an exact proof, see Strasser (1995)).

We therefore have a wide latitude in the construction of admissible tests. We will use our freedom of choice to construct tests which have additional nice properties, like the ease of implementation. To establish admissibility, it is enough to find a sequence of alternatives for which our test is equivalent to the Likelihood Ratio test. For these alternatives, our test will be optimal.

The null hypothesis for a given θ is denoted as

$$H_0(\theta) : \theta_t = \theta$$

and the sequence of local alternatives is given by

$$H_{1T}(\theta) : \theta_t = \theta + \frac{\eta_t}{\sqrt[4]{T}}. \tag{3.2}$$

Let Q_T^β denote the joint distribution of $(\eta_1, ..., \eta_T)$, indexed by the unknown parameter β. Let $P_{\theta,\beta}$ be the probability measure on $y_1, y_2, ..., y_T$ corresponding to $H_{1T}(\theta)$, and P_θ be the probability measure on $y_1, y_2, ..., y_T$ corresponding to $H_0(\theta)$. The ratio of the densities under $H_0(\theta)$ and $H_{1T}(\theta)$ is given by

$$\ell_T^\beta(\theta) \equiv \frac{dP_{\theta,\beta}}{dP_\theta} = \int \prod_{t=1}^T f_t\left(\theta + \eta_t/T^{1/4}\right) dQ_T^\beta / \prod_{t=1}^T f_t(\theta).$$

Theorem 3.1. *Under Assumptions 1-3, we have under $H_0(\theta)$*

$$\ell_T^\beta(\theta) / \exp\left(\frac{1}{2\sqrt{T}} \sum_{t=1}^T \mu_{2,t}(\beta,\theta) - \frac{1}{8} E\left(\mu_{2,t}(\beta,\theta)^2\right)\right) \xrightarrow{P} 1. \tag{3.3}$$

where the convergence in probability is uniform over β and $\theta \in \mathcal{N}$.

We can easily see from (2.3) that $\mu_{2,t}(\beta,\theta_0)$ is a stationary and ergodic martingale difference sequence, hence the central limit theorem applies. Moreover, for each sequence

$$\mathcal{N} \ni \theta_T \to \theta_0 \in \mathcal{N}, \tag{3.4}$$

the distribution of $\frac{1}{2\sqrt{T}} \sum_{t=1}^T \mu_{2,t}(\beta,\theta_T)$ will converge in distribution, under $H_0(\theta_T)$, to a Gaussian random variable with expectation 0 and variance $\frac{1}{4} E\mu_{2,t}(\beta,\theta_0)^2$.

Corollary 3.2. *For every sequence θ_T satisfying (3.4) and any β, the $P_{\theta_T,\beta}$ are contiguous with respect to P_{θ_T}.*

This result follows immediately from the CLT mentioned above and Strasser (1995). Denote

$$\ell_T\left(\theta_0 + \frac{1}{\sqrt{T}} d\right) \equiv \frac{dP_{\theta_0 + \frac{1}{\sqrt{T}} d}}{dP_{\theta_0}} = \frac{\prod_{t=1}^T f_t\left(\theta_0 + d/\sqrt{T}\right)}{\prod_{t=1}^T f_t(\theta_0)} = \exp\left\{\sum_{t=1}^T \left(l_t\left(\theta_0 + d/\sqrt{T}\right) - l_t(\theta_0)\right)\right\}.$$

Using a Taylor expansion around $\theta_0 + \frac{1}{\sqrt{T}} d$, we obtain the following lemma.

Lemma 3.3. *For all $\theta_0 \in \mathcal{N}$, and for all vectors d*

$$\ell_T\left(\theta_0+\frac{1}{\sqrt{T}}d\right)/\exp\left(-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d'l_t^{(1)}\left(\theta_0+\frac{1}{\sqrt{T}}d\right)+\frac{1}{2}E\left(d'l_t^{(1)}\left(\theta_0+\frac{1}{\sqrt{T}}d\right)\right)^2\right)\to 1 \tag{3.5}$$

uniformly (in d on all compacts) in probability.

Again, our regularity conditions guarantee the convergence of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d'l_t^{(1)}(\theta_0)$ to a normal distribution with variance $E\left(d'l_t^{(1)}(\theta_0)\right)^2$, hence again we can conclude that $P_{\theta_0+\frac{1}{\sqrt{T}}d}$ are contiguous with respect to P_{θ_0}. Since contiguity is a transitive relationship, we may conclude that for all vectors d, $P_{\theta_0+\frac{1}{\sqrt{T}}d,\beta}$ is contiguous with respect to P_{θ_0}. From

$$\frac{dP_{\theta_T,\beta}}{dP_{\theta_0}}=\frac{dP_{\theta_T,\beta}}{dP_{\theta_T}}\frac{dP_{\theta_T}}{dP_{\theta_0}}$$

we can conclude that with

$$\theta_T=\theta_0+\frac{1}{\sqrt{T}}d, \tag{3.6}$$

$$\begin{aligned}&\frac{dP_{\theta_T,\beta}}{dP_{\theta_0}}/\\&\left\{\exp\left(\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta_T)-\frac{1}{8}E\left(\mu_{2,t}(\beta,\theta_T)^2\right)\right)\exp\left(-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d'l_t^{(1)}(\theta_T)+\frac{1}{2}E\left(\left(d'l_t^{(1)}(\theta_T)\right)^2\right)\right)\right.\\\to\;&1\end{aligned}$$

where the convergence is - again - uniform in probability with respect to P_{θ_0}.

We now can proceed to construct optimal tests of $H_0(\theta_0)$ against the alternatives $H_{1T}(\theta_T)$. Fist assume that we know $\theta_0 \in \Theta$. Then contiguous alternatives to $H_0(\theta_0)$ are described by the probability measures

$$P_{\theta_T,\beta}, \tag{3.7}$$

where θ_T is given by (3.6). We now want to compare tests with respect to their power against these alternatives. In particular, we want to characterize tests by optimality properties. We want to start with a sequence of tests ψ_T and then show that there does not exist another sequence of tests φ_T which is asymptotically "better" for the null and all the contiguous alternatives. So let us formally define "better" tests.

Definition 3.4. *A sequence φ_T of tests is asymptotically better than ψ_T at θ_0 if it is "better" on the null*

$$\limsup\int\varphi_T dP_{\theta_0}\le\liminf\int\psi_T dP_{\theta_0} \tag{3.8}$$

and "better" on the alternatives, that is, for all θ_T and β

$$\liminf\int\varphi_T dP_{\theta_T,\beta}\ge\limsup\int\psi_T dP_{\theta_T,\beta}. \tag{3.9}$$

This definition is essentially the same as used by Andrews and Ploberger (1994) and a bit different from the one in Strasser (1995). Although the latter can be very useful when analyzing the asymptotic behavior of possible power functions for testing problems, our definition here proved more practical in econometric analysis because it directly deals with the asymptotic behavior of tests. Our definition here is, however, close enough to the one in Strasser (1995) so that we can use the standard proofs of optimality.

Definition 3.5. *A test ψ_T is said to be admissible if there exists no asymptotically better test.*

Let φ_T be some test statistics that has asymptotic level α (i.e. $lim \int \varphi_T dP_{\theta_0} = \alpha$) and asymptotic power function (i.e $\lim \int \varphi_T dP_{\theta_T,\beta}$ exists). Let $K \geq 0$ be an arbitrary constant, and ν be an arbitrary, but finite measure concentrated on a compact subset of $B \times \mathbf{R}^k$. Without limitation of generality we can assume that $\nu\left(B \times \mathbf{R}^k\right) = 1$. Then let us define the loss function

$$L\left(\varphi_T\right) = K \int \varphi_T dP_{\theta_0} - \int \left(\int \varphi_T dP_{\theta_0 + d/\sqrt{T},\beta}\right) d\nu(\beta, d). \tag{3.10}$$

By Fubini's theorem, we have

$$\begin{aligned} L\left(\varphi_T\right) &= \int (K - \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}})\varphi_T dP_{\theta_0} d\nu(\beta, d) = & (3.11)\\ & \int (K - \left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\})\varphi_T dP_{\theta_0} & (3.12)\end{aligned}$$

From (3.11) we can easily see that, for fixed K, $L(\varphi_T)$ is minimized by the tests ψ_T, which satisfy

$$\psi_T = \left\{\begin{array}{l} 1 \text{ if } \left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\} > K \\ 0 \text{ if } \left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\} < K \end{array}\right\}. \tag{3.13}$$

So the minimal loss only depends on the **distributions** of the $\left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\}$. We can easily see that the measures $\int P_{\theta_0 + d/\sqrt{T},\beta} d\nu(\beta, d)$ are contiguous with respect to P_{θ_0}, too. Hence the minimal loss equals

$$-\int (\left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\} - K)^{(+)} dP_{\theta_0}, \tag{3.14}$$

where, for an arbitrary real number x, $x^{(+)}$ denotes the positive part of x.

Let us now assume that we have a competing sequence of tests φ_T. Note that (3.13) does not uniquely determine a test: We do not care about the behavior of the test on the event $\left[\left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\} = K\right]$. Hence the following definition will be useful:

Definition 3.6. *The tests φ_T and φ'_T are asymptotically equivalent (with respect to our loss function L) if and only if for all $\varepsilon > 0$*

$$\lim E_{\theta_0} \left|\varphi_T - \varphi'_T\right| I \left[\left|\left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\} - K\right| > \varepsilon\right] = 0. \tag{3.15}$$

So, heuristically speaking, φ_T and φ'_T give us the same decision provided the test statistic $\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)$ is different from the critical value K. Moreover, we have the following result.

Theorem 3.7. *Suppose φ_T and ψ_T are asymptotically equivalent, where ψ_T is defined by (3.13). Then*

$$\lim \left(L(\psi_T) - L\left(\varphi_T\right)\right) = 0. \tag{3.16}$$

If φ_T and ψ_T are not asymptotically equivalent (in the above sense), then

$$\liminf \left(L(\psi_T) - L\left(\varphi_T\right)\right) < 0. \tag{3.17}$$

Hence (3.16) implies that ψ_T and φ_T are asymptotically equivalent.

Proof. We can easily see that $L(\psi_T) - L\left(\varphi_T\right) = \int (K - \left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\}) \left(\psi_T - \varphi_T\right) dP_{\theta_0}$. The construction of ψ_T and the fact that $0 \leq \varphi_T \leq 1$ imply that the integrand is nonpositive. Let $\varepsilon > 0$ be arbitrary. Let us define

$$r = (K - \left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\}). \tag{3.18}$$

Then

$$L(\psi_T) - L\left(\varphi_T\right) = \int rI\left[|r| > \varepsilon\right]\left(\psi_T - \varphi_T\right) dP_{\theta_0} + \int rI\left[|r| \leq \varepsilon\right]\left(\psi_T - \varphi_T\right) dP_{\theta_0}. \tag{3.19}$$

Since $\left|\psi_T - \varphi_T\right| \leq 1$, we have

$$\left|\int rI\left[|r| \leq \varepsilon\right]\left(\psi_T - \varphi_T\right) dP_{\theta_0}\right| \leq \varepsilon. \tag{3.20}$$

For asymptotically equivalent tests, $\int rI\left[|r| > \varepsilon\right]\left(\psi_T - \varphi_T\right) dP_{\theta_0} \to 0$, which proves (3.16). For (3.17), observe that if φ_T and ψ_T are not asymptotically equivalent, then there exists an $\eta > 0$ so that

$$\limsup E_{\theta_0} \left|\varphi_T - \psi_T\right| I \left[\left|\left\{\int \frac{dP_{\theta_0 + d/\sqrt{T},\beta}}{dP_{\theta_0}} d\nu(\beta, d)\right\} - K\right| > \eta\right] > 0 \tag{3.21}$$

The construction of ψ_T guarantees that $r\left(\psi_T - \varphi_T\right) \leq 0$. Hence $-\left|\varphi_T - \psi_T\right| rI\left[|r| > \varepsilon\right] = rI\left[|r| > \varepsilon\right]\left(\psi_T - \varphi_T\right) \leq rI\left[|r| > \eta\right]\left(\psi_T - \varphi_T\right)$ if $\eta \geq \varepsilon$, hence for all ε small enough lim inf

$\int rI\left[|r|>\varepsilon\right]\left(\psi_T-\varphi_T\right)dP_{\theta_0} < -\limsup E_{\theta_0}\left|\varphi_T-\psi_T\right| I\left[\left|\left\{\int \frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}}d\nu(\beta,d)\right\}-K\right|>\eta\right]$,
and together with (3.20) this proves our theorem. ■

We now can conclude from the above theorem that the tests ψ_T and all asymptotically equivalent sequences of tests are admissible. Any tests with genuine better power functions would have smaller loss, which is impossible. Hence we have to show that the our test is asymptotically equivalent to tests ψ_T.

For this purpose, let us first observe that the processes

$$Z_T\left(\beta,\theta\right)=\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}\left(\beta,\theta\right)-\frac{1}{8}E\left(\mu_{2,t}\left(\beta,\theta\right)^2\right)-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d'l_t^{(1)}\left(\theta\right)+\frac{1}{2}E\left(\left(d'l_t^{(1)}\left(\theta\right)\right)^2\right), \tag{3.22}$$

are, for all θ so that $\|\theta-\theta_0\| = O(1)/\sqrt{T}$ (and hence in particular the θ_T defined by (3.6)), uniformly tight in the space $C\left(B\right)$, the space of continuous functions on B. Indeed, since the $\mu_{2,t}\left(\beta,\theta_T\right)$ are stationary martingale differences, we can apply a central limit theorem and conclude that the $Z_T\left(\beta\right)$ converges in distribution (with respect to P_{θ_T}) to a Gaussian process with a.s. continuous trajectories. Since the P_{θ_T} are contiguous to P_{θ_0}, the limiting process(es) under P_{θ_0} must have continuous trajectories too, and we have uniform tightness of the distributions with respect to P_{θ_0}.

We now want to show that the tests ψ_T and the tests based on

$$\int \exp(Z_T(\beta,\theta_T))d\nu(\beta,d) \tag{3.23}$$

are asymptotically equivalent. We can easily see that a sufficient condition for asymptotic equivalence would be

$$\int \exp(Z_T(\beta,\theta_T))d\nu(\beta,d)/\int \frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}}d\nu(\beta,d)\to 1. \tag{3.24}$$

We know that for all finite sets β_i, d_i

$$\exp(Z_T(\beta_i,\theta_0+d_i/\sqrt{T}))/\frac{dP_{\theta_0+d_i/\sqrt{T},\beta_i}}{dP_{\theta_0}}\to 1. \tag{3.25}$$

So suppose that for all $\varepsilon>0$ and $\eta>0$ we could find a partition $S_1,...,S_K$ so that with probability greater than $1-\varepsilon$ for all i, $\left(\beta,d\right),\left(\gamma,e\right)\in S_i$ $\left|Z_T(\beta,\theta_0+d/\sqrt{T})-Z_T\left(\gamma,\theta_0+e/\sqrt{T}\right)\right|<\eta$, $\left|\frac{dP_{\theta_0+d/\sqrt{T},\beta}}{dP_{\theta_0}}-\frac{dP_{\theta_0+e/\sqrt{T},\gamma}}{dP_{\theta_0}}\right|<\eta$: Then (3.24) will be an easy consequence of (3.25).

The existence of such a partition for the Z_T is an immediate consequence of the uniform tightness of the distribution of Z_T. According to our assumptions, the difference between the Z_T and the log of the densities $\frac{dP_{\theta_0+d_i/\sqrt{T},\beta_i}}{dP_{\theta_0}}$ converges to zero uniformly in probability. Hence the density process is uniformly tight, too, which immediately guarantees the existence of the partition.

Let the tests ϕ_T reject when $\int\exp(Z_T(\beta,\theta_T))d\nu(\beta,d)>K$ and accept when $\int\exp(Z_T(\beta,\theta_T))d\nu(\beta,d)<K$. Then these tests are asymptotically equivalent to the tests ψ_T. Consequently, we have the following result:

Theorem 3.8. *Let φ_T be a sequence of tests that is asymptotically better (in the sense of definition 3.4) than ϕ_T . Then φ_T is asymptotically equivalent to ϕ_T .*

Proof. We just have shown that the ϕ_T are equivalent to the ψ_T, hence

$$\lim \left(L(\phi_T) - L(\psi_T)\right) = 0. \tag{3.26}$$

Since ψ_T are the tests with minimal loss function, we also have

$$\liminf \left(L(\varphi_T) - L(\phi_T)\right) \geq 0. \tag{3.27}$$

If δ is an arbitrary, finite measure and h_n measurable functions with $|h_n| \leq M$ for some M, then it is an easy consequence of Fatou's lemma that $\int \liminf h_n d\delta \leq \liminf \int h_n d\delta$. The definition 3.4 guarantees that $\liminf \left(\int \varphi_T dP_{\theta_T,\beta} - \int \phi_T dP_{\theta_T,\beta}\right) \geq 0$ and $\limsup \left(\int \varphi_T dP_{\theta_0} - \int \phi_T dP_{\theta_0}\right) \leq 0$. Since $L(\varphi_T) - L(\phi_T) =$ $K\left(\int \varphi_T dP_{\theta_0} - \int \phi_T dP_{\theta_0}\right) - \int \left(\left(\int \varphi_T dP_{\theta_0 + d/\sqrt{T},\beta}\right) - \left(\int \phi_T dP_{\theta_0 + d/\sqrt{T},\beta}\right)\right) d\nu(\beta, d)$, we can conclude that

$$\limsup \left(L(\varphi_T) - L(\phi_T)\right) \leq 0. \tag{3.28}$$

(3.27) and (3.28) allow us to conclude that $\lim \left(L(\varphi_T) - L(\phi_T)\right) = 0$, hence (3.26) also implies that $\lim \left(L(\varphi_T) - L(\psi_T)\right) = 0$. Then theorem 3.7 implies that φ_T and ψ_T are asymptotically equivalent. Since we did show that the ϕ_T are equivalent to the ψ_T, we have proved the theorem. ■

We now are able to construct asymptotically optimal tests for each parameter θ_0. The problem, however, is that we do not know θ_0. Hence we will try to find for each θ_0 a measure ν_{θ_0} so that the corresponding test statistic

$$\int \exp(Z_T(\beta, d)) d\nu_{\theta_0}(\beta, d) \tag{3.29}$$

does not depend on θ_0. For this purpose, define

$$d(\beta) = d(\beta, \theta_0) = \left(I(\theta_0)\right)^{-1} cov\left(\frac{1}{2}\mu_{2,t}(\beta, \theta_0), l_t^{(1)}(\theta_0)\right) \tag{3.30}$$

where $I(\theta_0)$ denote the information matrix. Then we have the following result:

Theorem 3.9. *Assume that J is a measure with mass 1 concentrated on a compact subset of B. Let d be as in (3.30), then define*

$$ST(\theta) = \int \left(\exp(Z_T(\beta,\ \theta + d(\beta,\theta)/\sqrt{T}))\right) dJ(\beta). \tag{3.31}$$

Let $\widehat{\theta}$ be the maximum likelihood estimator for θ under H_0,i.e.

$$\widehat{\theta} = \arg\max \sum l_t(\theta). \tag{3.32}$$

Then

$$expTS - ST(\theta_0) \to 0 \tag{3.33}$$

in probability under P_{θ_0},where

$$expTS = \int \left(\exp(TS_T(\beta, \widehat{\theta})\right) dJ(\beta), \tag{3.34}$$

and

$$TS_T(\beta, \widehat{\theta}) = \frac{1}{2\sqrt{T}} \sum \mu_{2,t}\left(\beta, \widehat{\theta}\right) - \frac{1}{2T}\widehat{\varepsilon}(\beta)' \widehat{\varepsilon}(\beta), \tag{3.35}$$

where $\widehat{\varepsilon}(\beta)$ is the residual from the OLS regression of $\frac{1}{2}\mu_{2,t}\left(\beta, \hat{\theta}\right)$ on $l_t^{(1)}\left(\hat{\theta}\right)$.

Let $P_{\widehat{\theta}}$ be the probability measure corresponding to the value *of the maximum likelihood estimator. (We can understand our parametric family as a mapping, which attaches to every θ a measure P_θ. Then the measure $P_{\widehat{\theta}}$ results from an evaluation of this mapping at $\widehat{\theta}$: It is a random measure). Let $K(\widehat{\theta})$ be real numbers so that*

$$P_{\widehat{\theta}}\left(\left[expTS < K(\widehat{\theta})\right]\right) \leq 1 - \alpha \tag{3.36}$$

$$P_{\widehat{\theta}}\left(\left[expTS > K(\widehat{\theta})\right]\right) \leq \alpha \tag{3.37}$$

and assume $K(\widehat{\theta}) \to K$. Then the tests φ_T, which reject if expTS $> K(\widehat{\theta})$, and accept if expTS $< K(\widehat{\theta})$, are for all θ_0 asymptotically equivalent under P_{θ_0} to tests rejecting if $ST(\theta_0) > K$, and accepting if $ST(\theta_0) < K$. Moreover, we have

$$P_{\theta_0}\left([ST(\theta_0) < K]\right) \leq 1 - \alpha \tag{3.38}$$

and

$$P_{\theta_0}\left([ST(\theta_0) > K]\right) \leq \alpha \tag{3.39}$$

Hence any sequence of tests better than φ_T is asymptotically equivalent to φ_T with respect to the probability measures P_{θ_0} for all $\theta_0 \in \Theta$.

The distribution of the $TS_T(\beta, \widehat{\theta})$ itself is of considerable interest, too. We are interested in functionals of $TS_T(\beta, \widehat{\theta})$, so we have to consider the limiting behavior of the whole function depending on the parameter β. Again, we restrict ourselves to compact subsets of B. Hence the appropriate limiting theory to consider is the convergence of distribution of random elements with values in the space of continuous functions defined on a compact subset of B.

Lemma 3.10. *Assume Assumptions 1 to 4 hold. Under H_0 and H_{1T}, we have*

$$TS_T(\beta, \widehat{\theta}) - \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\mu_{2,t}(\beta, \theta_0)}{2} - d(\beta)' l_t^{(1)}(\theta_0)\right) - \frac{1}{2} E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta, \theta_0)}{2} - d(\beta)' l_t^{(1)}(\theta_0)\right)^2\right)\right) \to 0 \tag{3.40}$$

uniformly on all compact sets. Moreover under H_0, we have

$$TS_T(\beta,\widehat{\theta}) \overset{D}{\Rightarrow} G(\beta),$$

where $\overset{D}{\Rightarrow}$ denotes the convergence in distribution of a sequence of stochastic processes and $G(\beta)$ is a Gaussian process with mean -$\frac{1}{2}E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta,\theta_0)}{2}-d(\beta)' l_t^{(1)}(\theta_0)\right)^2\right)$ and covariance

$$\begin{aligned} Cov(G(\beta_1),G(\beta_2)) &= E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta_1,\theta_0)}{2}-d(\beta_1)' l_t^{(1)}(\theta_0)\right)\left(\frac{\mu_{2,t}(\beta_2,\theta_0)}{2}-d(\beta_2)' l_t^{(1)}(\theta_0)\right)\right) \\ &\equiv k(\beta_1,\beta_2). \end{aligned}$$

Under H_{1T}, $TS_T(\beta,\widehat{\theta})$ converges to a Gaussian process with mean $k(\beta,\beta_0)-\frac{1}{2}k(\beta,\beta)$ and variance $k(\beta_1,\beta_2)$, where β_0 is the true value of the parameter β under the alternative.

The last statement follows from Le Cam's third lemma (see van der Vaart, 1998) and from the fact that the joint distribution of the $TS_T(\beta,\widehat{\theta})$ and the logarithms of the densities of the local alternatives converges to a joint normal, and these two Gaussian random variables are correlated.. With the help of this lemma, we can conclude that our test has nontrivial power against local alternatives if $E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta,\theta_0)}{2}-d(\beta)' l_t^{(1)}(\theta_0)\right)^2\right)>0$.

It is, however, also possible that

$$E_{\theta_0}\left(\left(\frac{\mu_{2,t}(\beta,\theta_0)}{2}-d(\beta)' l_t^{(1)}(\theta_0)\right)^2\right)=0. \tag{3.41}$$

This case is not that implausible. Indeed we have

$$\begin{aligned} & E_{\theta_0}\left(\left(\frac{\mu_{2,t}}{2}-d' l_t^{(1)}\right)^2\right) \\ =\ & E_{\theta_0}\left(\left(\frac{\mu_{2,t}}{2}\right)^2\right)-2d' E_{\theta_0}\left(l_t^{(1)}\frac{\mu_{2,t}}{2}\right)+d'\left(I(\theta_0)\right)^{-1}d \\ =\ & E_{\theta_0}\left(\left(\frac{\mu_{2,t}}{2}\right)^2\right)-E_{\theta_0}\left(l_t^{(1)}\frac{\mu_{2,t}}{2}\right)'\left(E_{\theta_0}\left(l_t^{(1)}l_t^{(1)\prime}\right)\right)^{-1}E_{\theta_0}\left(l_t^{(1)}\frac{\mu_{2,t}}{2}\right) \end{aligned}$$

using (3.30). Hence (3.41) is satisfied if and only if $\mu_{2,t}$ belongs to the linear span of the components of $l_t^{(1)}$. Assume for a moment that $\rho=0$ and all the other prerequisites of Assumptions 3 and 4 are fulfilled. Then $\mu_{2,t}$ is a linear functional of the second-order derivatives of the log-likelihood, namely $h'\left(\frac{\partial^2 l_t}{\partial\theta\partial\theta'}+\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_t}{\partial\theta}\right)'\right)h$. Then (3.41) means that the second order derivatives can be written as a linear combination of the scores. This is a geometric condition, which has profound statistical implications: E.g. in Murray and Rice (1993), p. 16 it is used to characterize linear exponential families. A typical example would be the normal distribution. We have two parameters, mean and variance, and we can easily see that if we take $h = (1, 0)'$ (our first parameter should be the mean) (3.41) is fulfilled. This corresponds to testing for independent mixture of two normals with different unknown means and same unknown variance.

If (3.41) is fulfilled, then it is impossible to construct a test with nontrivial power against these specific local alternatives. The $TS_T(\beta, \widehat{\theta})$ are consistent approximations of the log-density of one measure under the null (corresponding to θ_0 and to $\theta_0 + d/\sqrt{T}$, β, respectively). If the density between these two measures converges to 1, then any reasonable distance like e.g. total variation converges to zero. So in this kind of situation null and alternative are not distinct probability measures, which makes it impossible to construct consistent tests. Any test will have trivial local power for an alternative in $T^{-1/4}$. However our test may have non trivial power against a local alternative of order $T^{-1/6}$ for instance. This means that our test may still have power against a **fixed** alternative.

Moreover, under Assumption 3, this phenomenon is the exception rather than the rule. The following proposition characterizes the set of alternatives against which our test does not have local power.

Proposition 3.11. *Suppose Assumptions 1 to 4 hold. Assume furthermore that for all* t, s, $h'\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)' h$ can not *be represented as a linear combination of components of* $\left(\frac{\partial l_t}{\partial\theta}\right)$. *Then for each* h, *there exist at most finitely many* ρ *so that (3.41) is fulfilled.*

Proof. First of all let us observe that

$$\mu_{2,t}(\beta,\theta) = c^2 h'\left[\left(\frac{\partial^2 l_t}{\partial\theta\partial\theta'} + \left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_t}{\partial\theta}\right)'\right) + 2\sum_{s<t}\rho^{(t-s)}\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)'\right] h$$

Let us assume that for one h there exist infinitely many values of ρ so that (3.41) is fulfilled. We can easily see that $\mu_{2,t}(\beta,\theta)$, and hence d, are analytic functions of ρ. Therefore $E_{\theta_0}\left(\left(\frac{\mu_{2,T}(\beta,\theta_0)}{2} - d' l_t^{(1)}(\theta_0)\right)^2\right)$ must be an analytic function too. We did assume that this function has infinitely many zeros in a finite interval, hence it must be identically zero. Hence

$$c^2 h'\left[\left(\frac{\partial^2 l_t}{\partial\theta\partial\theta'} + \left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_t}{\partial\theta}\right)'\right) + 2\sum_{s<t}\rho^{(t-s)}\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)'\right] h = d(c,h,\rho)'\left(\frac{\partial l_t}{\partial\theta}\right)$$

for all ρ. Since both sides of the equation are analytic functions, their derivatives (with respect to ρ) must be also equal. Hence

$$2c^2 h'\left(\frac{\partial l_t}{\partial\theta}\right)\left(\frac{\partial l_s}{\partial\theta}\right)' h = d'_{t-s}\left(\frac{\partial l_t}{\partial\theta}\right),$$

where d'_{t-s} is the coefficient of $\rho^{(t-s-1)}$ in the derivative of $d(.,.,.)$ with respect to ρ. In the case where $c^2 \neq 0$, this contradicts our assumption. ■

The restriction to prior measures with compact support might be a bit restrictive. In most cases, we should be able to approximate prior measures with noncompact support by ones with compact support. In cases where (3.41) is fulfilled, we will, however, encounter a difficulty. For our test statistic, we have to compute expTS= $\int \left(\exp(TS_T(\beta,\widehat{\theta}))\right) dJ(\beta)$. For the values of β where (3.41) holds the corresponding $TS_T(\beta,\widehat{\theta})$ will converge to zero. It is, however, difficult to get uniform convergence. Hence we will not derive theorems for these measures here.

The admissibility of the sup test could be proved using a similar approach to Andrews and Ploberger (1995).

4. Monte Carlo study

We start with a very simple model with switching intercept and an uncorrelated and homoskedastic noise component,

$$y_t = \alpha_0 + \alpha_1 S_t + \omega_0 \epsilon_t$$

where

$$\begin{aligned} P(S_t &= 1 \mid S_t = 1) = p \\ P(S_t &= -1 \mid S_t = -1) = q \end{aligned}$$

and $\epsilon_t \sim iid\mathcal{N}(0,1)$. We compare our test with Garcia's (1998) likelihood ratio test. Garcia's test requires the estimation of the model under the null and the alternative and the problem of local maxima arises under the alternative (see Hamilton (1989) and Garcia and Perron (1996)). As a result, $1,000$ replications will only produce a fraction of positive log likelihood ratios, and among these a lot of values close to zero. Garcia circumvents this problem by using 12 sets of starting values for the optimization and by taking the maximum over the values obtained. We apply this method, which turns out to be quite successful.

To compare the power performances between the two tests, we use 1000 replications and 100 observations. We generate exactly the same data for both cases. Under $H_0, \alpha_0 = \alpha_1 = 0, \omega_0 = 1$. Under $H_1, \alpha_0 = 0, \alpha_1 = c/\sqrt[4]{T}, p = q = 0.75$ and $\omega_0 = 1$. We use our supTS test discussed in Section 2. We maximize over h and ρ with $\rho \in (-0.7, 0.7)$. Our test statistic is asymptotically equivalent to Garcia's in the sense that both are some kind of likelihood ratio tests and hence they are expected to have similar powers.

Figure 8.1 plots the size-corrected powers for various values of c. As expected, the patterns for both tests are similar. Our power is slightly higher than Garcia's in general, but is a tiny bit smaller for $c = 4$. Our test has the great advantage that it only requires estimating the parameters under the null. As a consequence, it is easy to program and execute. Moreover, we find that, the size-corrected power does not change much when maximizing our supTS over h and ρ by generating h uniformly over the unit sphere and

ρ selected from an equispaced grid. But it greatly saves time (about 1/4 time for above model).

Then we apply the supTS test to more general models. To find the maximum over h and ρ, we generate h uniformly over the unit sphere and ρ is selected from an equispaced grid of $(-0.7, 0.7)$. The number of values for h is 30 and that of ρ is 60. We obtain the empirical critical values with 1000 iterations and sample size is taken to be 100. Then we plot the size-corrected power with the same number of iterations and same sample sizes.

Linear model with an intercept term:

$$\begin{aligned} y_t &= x_t' \left(\beta + \frac{C\eta_t}{\sqrt[4]{T}} \right) + \varepsilon_t \\ \varepsilon_t &\sim iid\mathcal{N}(0,1) \end{aligned}$$

$\beta = (1,1)'$, $C = (c_1, c_2)'$. $x_t = (1, x_{1t})'$ with x_{1t} ∼iid$\mathcal{N}(3, 400)$. η_t is a two-State Markov chain that takes the values 1 and -1 with transition probabilities $P\left(\eta_t = 1|\eta_{t-1} = 1\right) = 0.75$ and $P\left(\eta_t = -1|\eta_{t-1} = -1\right) = 0.75$.

In the simulations, we set $c_1 = c_2 = c$ and vary them. The size-corrected power as a function of c is plotted in Figure 8.2.

ARCH(1) model:

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= (\frac{1}{4} + \frac{c_1\eta_t}{\sqrt[4]{T}}) + (\frac{1}{4} + \frac{c_2\eta_t}{\sqrt[4]{T}}) y_{t-1}^2 \\ \varepsilon_t &\sim iid\mathcal{N}(0,1) \end{aligned}$$

η_t is a two-State Markov chain that takes the values 1 and -1 with transition probabilities $P\left(\eta_t = 1|\eta_{t-1} = 1\right) = 0.75$ and $P\left(\eta_t = -1|\eta_{t-1} = -1\right) = 0.75$. The size-corrected power is shown in Figure 8.3 as a function of $c = c_1 = c_2$.

IGARCH(1,1):

The model is as follows:

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= (\frac{1}{2} + \frac{c_1\eta_t}{\sqrt[4]{T}}) + (\frac{1}{2} + \frac{c_2\eta_t}{\sqrt[4]{T}}) \sigma_{t-1}^2 + (\frac{1}{2} + \frac{c_3\eta_t}{\sqrt[4]{T}}) y_{t-1}^2 \\ \varepsilon_t &\sim iid\mathcal{N}(0,1) \end{aligned}$$

Note that $\alpha_1 + \beta_1 = 1$. Here, we let η_t take the values 0 and -1 with transition probabilities $P\left(\eta_t = 0|\eta_{t-1} = 0\right) = 0.75$ and $P\left(\eta_t = -1|\eta_{t-1} = -1\right) = 0.75$. c_1, c_2, and c_3 are taken to be equal. See size-corrected power in Figure 8.4.

This simulation study shows that our test has satisfactory power in small samples.

5. A Markov-switching model for explosive bubbles

Let P_t and D_t be the stock price and dividend at time t. $0 < (1+r)^{-1} < 1$ is the discount rate (assumed constant). The size of a bubble is the difference between P_t and the market fundamental price solution, F_t, (which equals the expected present value of future dividends)

$$B_t = P_t - F_t.$$

Rational expectation predicts that

$$B_t = (1+r)^{-1} E_t B_{t+1}.$$

Evans (1991) argues that an interesting class of rational bubbles have the property to collapse with probability one. He proposes an example of such a bubble:

$$\begin{aligned} B_{t+1} &= (1+r) B_t u_{t+1} \text{ if } B_t \leq \alpha, \\ B_{t+1} &= \left[\delta + \frac{(1+r)}{\pi}\theta_{t+1}\left(B_t + \frac{\delta}{1+r}\right)\right] u_{t+1} \text{ if } B_t > \alpha, \end{aligned} \tag{5.1}$$

where u_{t+1} is exogenous iid with $E_t u_{t+1} = 1$ and θ_{t+1} is exogenous, iid $\mathcal{B}(1,\pi), 0 < \pi \leq 1$. The dynamic of B_t in (5.1) is partly threshold, partly mixture. This model was meant by Evans as illustrative only. However, it is interesting because it suggests that the price deviations from the fundamental variable may explode and shrink periodically while being consistent with the rational expectation assumption. To test this idea, we proceed in two steps.

First we estimate the following cointegration relationship between $\ln(P_t)$ and $\ln(D_t)$

$$\ln(P_t) = \hat{a}_0 + \hat{a}_1 \ln(D_t) + y_t \tag{5.2}$$

by ordinary least-squares. As D_t plays the role of fundamentals (in the spirit of Lucas, 1978), we expect the residual y_t to behave as a periodically collapsing bubble. Then we fit on y_t the Markov-switching model:

$$\Delta y_t = \sum_{s_t} \alpha_{s_t} + \sum_{s_t} \beta_{s_t} y_{t-1} + \sum_{i=1}^{l} \sum_{s_t} \phi_{s_t i} \Delta y_{t-i} + \varepsilon_t \tag{5.3}$$

where $\varepsilon_t \sim iid\mathcal{N}(0,\sigma^2)$. S_t is an exogenous three-state Markov chain that takes the values 1,2, and 3 and has for transition probabilities $0 < p_{ij} < 1$. Because the labels of the regimes are interchangeable, we set $\beta_1 \geq \beta_2 \geq \beta_3$. The parameter of interest is $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \phi_{1i}, \phi_{2i}, \phi_{3i} : i = 1, .., l, p_{ij} : i, j = 1, 2, 3)'$.

We know from Yao and Attali (2000) that $\{y_t\}$ may be stationary even if there is an explosive root in one of the regimes. Therefore testing the stationarity of $\{y_t\}$ alone does not permit to conclude against the presence of bubbles.

Proposition 5.1. *Assume* $\ln(D_t)$ *is strictly exogenous for* ε_t*, in the sense that* ε_t *is uncorrelated with* $\ln(D_1), \ln(D_2), .., \ln(D_T)$*. The MLE estimates of* (a_0, a_1, θ) *coincide with the estimators obtained from a two-step procedure consisting in estimating* $(a_0, a_1)'$ *by OLS in (5.2) first and then applying MLE on (5.3). Moreover the resulting* $\hat{\theta}$ *are independent of* $(\hat{a}_0, \hat{a}_1)$ *implying that the first step does not affect the second step.*

Data

We use monthly US data from 1871-01 to 2002-06 (T =1578) for real stock prices and real dividends. All prices are in January 2000 dollars. These data are taken from Shiller's web site http://www.econ.yale.edu/˜shiller and described in Shiller (2000).

Results

Applying a BIC criterion on an autoregressive model reveals that 2 lags are best, hence we set $l = 1$ in Model (5.3). The augmented Dickey Fuller test rejects the null of a unit root on y_t at a 1% level. We apply the supTS test (described in Section 2) where the maximum over h and ρ is obtained by drawing h uniformly over the unit sphere (30 values used) and by taking the values of ρ in an equally spaced grid over $(-0.7, 0.7)$ (60 values used).

Empirical critical values are computed from 1000 iterations for a sample size of 1576. The values of the parameters used to simulate the series are those obtained when estimating the model under H_0. The critical values are 5.6577635, 4.2483499, 3.7680360 at 1%, 5% and 10% respectively. The test statistic for our data is 22.938546. Hence our linearity test rejects strongly the null of a linear model versus a Markov-switching alternative, suggesting that at least two regimes should be used to fit the data. We estimate model (5.3) by maximum likelihood using the EM algorithm described in Hamilton (1989). We use 12 sets of starting values and select the one corresponding to the largest value of the likelihood.

	estimate	**standard error**
α_1	-0.100	0.019
β_1	0.038	0.038
ϕ_1	-0.195	0.204
α_2	0.002	0.001
β_2	$-.010$	0.004
ϕ_2	0.321	0.033
α_3	0.057	0.039
β_3	-0.216	0.057
ϕ_3	1.431	0.115
σ	0.001	6.9e-5

The estimated transition matrix P with elements $p_{ij} = P(S_{t+1} = i | S_t = j)$ is given by

$$P = \begin{bmatrix} 0.253 & 0.023 & 0.146 \\ 0.232 & 0.973 & 0.735 \\ 0.515 & 0.004 & 0.119 \end{bmatrix}$$

and the estimated stationary distribution is $P(S_t = 1) = 0.034$, $P(S_t = 2) = 0.942$, and $P(S_t = 3) = 0.024$.

All the coefficients are significantly different from 0. Regime 1 ($S_t = 1$) corresponds to an explosive root with negative drift. In this regime, the trend (-0.1) dominates corresponding to declines of 10%. The second regime ($S_t = 2$) corresponds a near unit-root with a slight positive drift. In this regime the process is stationary because the null hypothesis $H_0 : \beta_2 = 0$ is rejected. 94% of the data lies in this regime, which is very persistent. Finally Regime 3 ($S_t = 3$) corresponds to a strong mean-reverting process. By filtering, we compute the probabilities to be in Regime 1 conditional on the data: $P(S_t = 1|y_1, ..., y_T)$. When $P(S_t = 1|y_1, ..., y_T) > 0.5$, it is considered that the process at date t is in Regime 1. The following months lie in Regime 1:

1873 (9-11), 1880 (4), 1893 (5-7), 1907 (3,8,10,11), 1917 (11), 1929 (10, 12), 1930 (5,6,10,12), 1931 (9,10,12), 1932 (4-6,10), 1933 (2), 1934 (5), 1937 (4,6,9,10), 1939 (4), 1940 (5), 1946 (9), 1950 (7), 1962 (5), 1970 (5), 1973 (11), 1974 (7,9), 1980 (3), 1981 (9), 1987 (10).

We recognize the big crashes such as October 1929 and October 1987. We can compare our results with those of Pagan and Sossounov (2003) on bull and bear markets. We see that Regime 1 identifies the month just preceding a trough of the US stock market cycles as reported in Pagan and Sossounov (1962/6, 1970/6, 1974/9, and 1987/11). It means that the periods of negative drift correspond to a crash (in other words, the burst of a bubble). The process spends most of the time in the near unit-root regime. This asymmetric pattern exhibiting slow increases and quick decreases is consistent with the presence of periodically collapsing bubbles.

Related literature

Diba and Grossman (1988) apply Dickey-Fuller test on the price and dividends both in level and first difference. They also test whether P_t and D_t are cointegrated. As they found that P_t and D_t are both integrated of order 1 and mutually cointegrated, they conclude that there is no bubble. Evans (1991) shows that Diba and Grossman just tested for the presence of a specific (linear) type of bubble. Since Evans (1991) pointed out the shortcomings of traditional unit-root tests to establish the presence of bubbles, there have been only a few attempts to devise a test. Van Norden and Schaffer (1993) and van Norden and Vigfusson (1996) use a mixture model where the probability of belonging to one regime depends on the lagged value of y_t. Hall, Psaradakis, and Sola (1999) use a Markov-switching model to model the consumer price index and exchange rate in Argentina (1983 to 1989) and find a bubble in the exchange rate in 1984-1985. Psaradakis, Sola, and Spagnolo (2001) apply a test of stochastic unit root on German hyperinflation data. Chirinko and Schaller (2001) apply an orthogonality test (GMM type) to show that there has been a bubble in Japanese equity market in the eighties.

Our data have been previously investigated for bubbles by Taylor and Peel (1998) and Bohl (2001). Both papers reject the presence of periodically collapsing bubbles. Taylor and Peel use a new test that is robust to the presence of skewness and kurtosis in the data. Bohl uses a MTAR model. The MTAR is a Threshold model where the change of regimes is triggered by the lagged value of Δy_t. From an economic point of view, the change of regime should be triggered by the lagged value of y_t and not Δy_t. This may be the reason why Bohl's test fails to support the bubble hypothesis.

6. Appendix A: Notations

6.1. Tensors

Central to the proofs in this paper are Taylor series expansions to the fourth order. We will have to organize and manipulate expressions involving multivariate derivatives of higher orders. We therefore will be careful with our notation. Clearly it would be possible to use partial derivatives, but then our expressions will get really complicated. Hence we will adopt some elements from multilinear algebra, which will facilitate our computations.

Key to our analysis is the concept of a multilinear form. Consider vector spaces V, F. Then a multilinear form of order p from V into F is a mapping M from $V \times .. \times V$ (where we take the product p times) to F which is linear in each of the arguments. So

$$\begin{aligned} & \lambda M(x^{(1)}, x^{(2)}, ..., x_1^{(i)}, ..., x^{(p)}) + \mu M(x^{(1)}, x^{(2)}, ..., x_2^{(i)}, ..., x^{(p)}) && (6.1) \\ = \; & M(x^{(1)}, x^{(2)}, ..., \lambda x_1^{(i)} + \mu x_2^{(i)}, ..., x^{(p)}). && (6.2) \end{aligned}$$

We will use only an elementary form of tensors (only rather straightforward fields, defined on parts of $\mathbf{R}^k$). Here we give a brief introduction to the necessary formalism.

The first important concept we need to discuss is the definition of a derivative. Essentially, we will follow the differential calculus outlined in Lang (1993), p. 331 ff. Let f be a function defined on an open set O of the finite-dimensional vector space V into the finite dimensional space F. Then f is said to be differentiable if for all $x \in O$ there exists a linear mapping $Df = Df(x)$ from V to F so that

$$\lim_{r \to 0} \sup_{\|h\|=r} \|f(x+h) - f(x) - Df(x)(h)\| / r \to 0. \tag{6.3}$$

The above expression should not be misinterpreted. $Df(x)$ attaches to each $x \in O$ a linear mapping, so $Df(x)(h)$ is for each $h \in V$ an element of F. $Df(x)$ is called a Frechet-derivative. It is in a way a formalization of the well known "differential" in elementary calculus. So $Df(x)$ is a linear mapping between V and F. It is an elementary task to show that the space of all linear mappings between V and F, denoted by $L(V, F)$ is a finite dimensional vector space again. Hence we can consider the mapping

$$x \to Df(x), \tag{6.4}$$

which maps O into $L(V, F)$, so we may use the concept of Frechet-differentiability again and differentiate Df. We then get the second derivative $D^2f(x)$. This second derivative at a point is a linear mapping from V to $L(V, F)$ (an element from $L(V, L(V, F))$). That means that, for each $h \in V$, $D^2f(x)(h)$ is an element of $L(V, F)$, so for $k \in V$ $D^2f(x)(h)(k)$

is an element of F. Moreover, we can easily see that - by construction - the expression $D^2 f(x)(h)(k)$ is linear in h and k. Hence $D^2 f(x)$ maps each pair (h, k) into F and is linear in each of the arguments, so we can think of $D^2 f(x)$ as a bilinear form from $V \times V$ into F.

It is easily seen that, in case f has enough "derivatives", we can iterate this process and define the n-th derivative $D^n f$ as derivative of $D^{n-1} f$,

$$D^n f = D(D^{n-1} f). \tag{6.5}$$

Again we can interpret $D^n f$ as an element of $L(V, L(V, ...L(V, F))))$ or - again - as a multilinear mapping from $V \times V \times V \times V.. \times V$ into F. This means that $D^n f\ (x)$ attaches to each n-tupel $(x_1, ..., x_n)$ of elements of V an element of F, in such a way that the mapping is linear in each of its arguments.

Most importantly, we have again a Taylor formula

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2} D^2 f(x)(h, h) + ... + \frac{1}{n!} D^n f(x)(h, ..h) + R_n \tag{6.6}$$

with

$$R_n = \frac{1}{n!} \int_0^1 (1-t)^n D^{n+1} f(x+th)(h, ..., h) dt, \tag{6.7}$$

if f is at least $n+1$ times continuously differentiable.

Furthermore it is relatively easy to verify that f being n times continuously differentiable

$$D^n f \text{ is symmetric} \tag{6.8}$$

i.e.

$$D^n f(x)(h_1, ..., h_n) = D^n f(x)(h_{\pi(1)}, ..., h_{\pi(n)}) \tag{6.9}$$

for every permutation π.

Moreover, let us consider for fixed x, h the function $g(t) = f(x+ht)$ for t in a neighborhood of 0, and let $g^{(n)}$ be the n-th derivative of g. Then

$$g^{(n)}(0) = D^n f(x)(h, ..., h). \tag{6.10}$$

It is now an elementary, but tedious, exercise to show that due to the symmetry (6.9) the multilinear form $D^n f(x)$ is uniquely defined by its values $D^n f(x)(h, ..., h)$. (As an example, it might be instructive to consider the case of a scalar bilinear form B: We can easily see that

$$B(h, k) + B(k, h) = \frac{1}{4} (B(h+k, h+k) - B((h-k, h-k)). \tag{6.11}$$

Symmetry implies that the left hand side of the above equation equals $2B(h, k) = 2B(k, h)$.)

This result allows us to "translate" all the well-known results from elementary calculus to our formalism. Clearly the derivative is linear, we have a product rule - if f and g are scalar functions, then $D(fg) = f \cdot Dg + (Df) \cdot g$, and more importantly we have a chain rule: If we compose functions f, g we have

$$D(f \circ g) = Df(Dg). \tag{6.12}$$

The algebra of multilinear forms is often treated as a special case of tensor algebra. Although this branch of mathematics is well developed, it is rarely used in econometrics. Furthermore, many of the advanced concepts are of no use to us. Hence we will introduce here a "baby" version of tensor algebra. In particular, we will sacrifice one of the basic ideas of modern analysis, namely that geometric concepts should be defined without reference to a coordinate system. Hence we will call our objects "baby tensors". The experts will see that they share many of the usual properties of tensors. The key simplification will be that we fix our reference space and the coordinate system once and for all - we simply forbid the use of other coordinate systems and spaces.

We are in a rather advantageous position:

- We are mostly interested in manipulating the derivatives of a scalar function, namely the logarithm of the likelihood function.
- Working independently of a coordinate system is not a priority for us (contrary to theoretical physics, where gauge invariance plays a major role).
- For us, tensors are tools to work with multilinear forms. Hence, most of our tensors are symmetric.

Assume that our reference, finite dimensional vector space V is $k-$dimensional and that $b_1, ..b_k$ is a basis for this space. Although the basis is arbitrary, we will from now on **assume this basis to be fixed**. It is **essential** for our approach that we **fix the underlying vector space and the basis**, since all of our definitions relate in one way or another to our chosen basis. It should be noted that we follow this approach not out of necessity - coordinate independent definitions of tensors are commonplace in differential geometry and mathematical physics, but purely out of convenience. E.g. we do not need to distinguish between co- and contravariant tensors - so we do not have to distinguish between "upper" and "lower" indices.

With the help of our basis, any vector x can uniquely be written as

$$x = \sum_{i=1}^{k} x_i b_i. \tag{6.13}$$

Let us now assume that M is a scalar multilinear form (i.e. the values of M are real numbers). Then, using linearity, we have

$$M(x^{(1)}, x^{(2)}, ..., x^{(p)}) = \sum M(b_{i_1}, ..., b_{i_p}) x^{(1)}_{i_1} x^{(2)}_{i_2} .. x^{(p)}_{i_p}, \tag{6.14}$$

where the sum symbol corresponds to p sums extending over all values of $i_1, ..., i_p$ between 1 and k. So we can easily see that there is a one-to-one correspondence between the k^p numbers $M(b_{i_1}, ..., b_{i_p})$ and the multilinear forms. For each set of numbers we define a uniquely determined multilinear form, and for each multilinear form we can find coefficients. Hence it seems sensible to give this scheme a name.

Definition 6.1. *A scheme of k^p numbers $\left(c_{i_1,...,i_p}\right)$, indexed by p-tuples $i_1, ..., i_p$, where each indice varies between 1 and k, will be called a* **baby tensor** *(abbreviated* **bt***) of order p. A* **baby tensor field** *(***btf***) is a mapping which attaches to each element of an open subset of our basic vector space a baby tensor. The coefficients $\left(c_{i_1,...,i_p}\right)$ are the bt's* **coordinates***. The case $p = 0$ is admissible: The corresponding bt is a constant and the btf is just a scalar function.*

This concept allows us to work with multilinear forms and related mathematical objects without having to discuss tensor algebra. Here we present a compromise. On the one hand, our objects should facilitate computations with multilinear forms, and on the other hand, the objects should not be too abstract. A bt of order one is simply a vector, a bt of order two is a $k \times k-$matrix.

1. Now let us discuss a few properties of baby tensors and baby tensor fields. We can see immediately that baby tensors and multilinear forms are essentially the **same** object: (6.14) demonstrates that there is an one-to-one correspondence between multilinear forms and bts. If we add two multilinear forms, we get another multilinear form whose bt equals the sum of the bts of the summands. If we multiply a multilinear form by a constant factor, we can easily see that the bt of the resulting linear form equals the product of the original bt and the factor in question. Hence we can see that this correspondence is **a linear** one-to-one mapping, an isomorphism. To simplify the notation, we will denote the multilinear form and the bt by the same symbol.

2. Let us call a bt $C = \left(c_{i_1,...,i_p}\right)$ **symmetrical** if and only for all $(i_1, ..., i_p)$ and all permutations π of numbers between 1 and k

$$c_{i_1,...,i_p} = c_{\pi(i_1),...,\pi(i_p)}. \tag{6.15}$$

   We can easily see that a bt is symmetrical if and only if the corresponding multilinear form is symmetrical (in the sense of (6.9)). For a bt $C = \left(c_{i_1,...,i_p}\right)$ define its symmetrization $C^{(S)}$ by

$$\left(C^{(S)}\right)_{i_1,...,i_p} = \frac{1}{k!} \sum_{\text{all permutation } \pi \text{ of } \{1,..,k\}} c_{\pi(i_1),...,\pi(i_p)}. \tag{6.16}$$

   Then $C^{(S)}$ is symmetrical. Moreover, for all $h \in V$

$$C(h, ...h) = C^{(S)}(h, ..., h), \tag{6.17}$$

   and, for any bt C, $C^{(S)}$ is the only symmetrical tensor with the property (6.17).

3. Another special case of multilinear forms are our derivatives of scalar functions defined on open subsets of our space V. Hence we can associate to a derivative a bt, too. We can easily see that the bt associated with $D^n f$ can be calculated in the following way. Define the function g by

$$g((x_1, ..., x_p) = f(\sum x_i b_i), \tag{6.18}$$

where the b_i are our fixed basis vectors. Then the corresponding bt is given by $\left(\frac{\partial^n g}{\partial x_{i1} \partial x_{i_2} .... \partial x_{i_n}}\right)_{(i_1,...,i_n)}$.

4. There is also another technique for computing $D^n f$, which we will use below. Define for fixed x and $h \in V$, the function

$$g_h(t) = f(x + th). \tag{6.19}$$

Then - following (6.10) - we can conclude that $D^n f(h, h, ..h) = g_h^{(n)}(0)$, where $g_h^{(n)}$ is the usual n-th derivative. Now suppose we can find a bt C so that for all h

$$C(h, ..., h) = g_h^{(n)}(0). \tag{6.20}$$

Then - due to (6.17) and symmetry of the derivative - we can conclude that $D^n f = C^{(S)}$.

5. Apart from the usual operations, we also can define the **tensor product** between bts. Let A and B be bts of order p and q with coordinates $\left(a_{(i_1,...,i_p)}\right)$ and $\left(b_{(i_1,...,i_q)}\right)$, respectively. Then the tensor product $A \otimes B$ is a bt of order $p + q$ with coordinates

$$a_{(i_1,...,i_p)} b_{(i_{p+1},...,i_{p+q})}. \tag{6.21}$$

Although the definition of the tensor product looks similar to the Kronecker product, these two concepts should not be confused. A Kronecker product of two matrices is again a matrix. In contrast, the tensor product of two bts of order two is a bt of order four. It is interesting to consider the properties of the corresponding multilinear forms:

$$(A \otimes B)(h_1, .., h_{p+q}) = A(h_1, ...h_p) B(h_{p+1}, ..., h_{p+q}). \tag{6.22}$$

The tensor product of symmetric tensors, however, is in general not symmetric.

6. Bts have a third interpretation. We can think of (6.14) not only as a definition for a multilinear form, but also as a definition for a linear form on the space of bts. Let T be a given bt of order p with coordinates $\left(t_{i_1,...,i_p}\right)$. We can define a mapping from the space of all bts of order p to the real numbers by the following rule: We attach to each bt C with coordinates $\left(c_{i_1,...,i_p}\right)$ the number

$$\sum t_{i_1,...,i_p} c_{i_1,...,i_p}. \tag{6.23}$$

We can immediately see that this mapping is an isomorphism. Again we denote the bt and this mapping by the same symbol. So for each bt C, $T(C)$ is a real number given by (6.23). This formalism is very useful when we want to compute expectations of multilinear forms. Suppose we have a bt T and we want to compute the multilinear form $T(h_1, ...h_p)$. Then we can see from (6.14),(6.23) that $T(h_1, ...h_p)$ equals $T(h_1 \otimes .. \otimes h_p)$ (interpreted as a linear mapping from the space of bts).

7. Now suppose $H_1, ..., H_p$ are random variables with values in our reference space V, and T is a bt. Suppose we want to compute the expectation of

$$T(H_1, ..., H_p). \tag{6.24}$$

Then we can use the above interpretation of bt as a linear mapping and conclude that the above expression equals $T(H_1 \otimes ... \otimes H_p)$. We should not have any conceptual problem with $H_1 \otimes ... \otimes H_p$: This is a random variable with values in the space of all bts of order p, which is a finite dimensional space. Moreover, since T is a linear mapping, we have

$$ET(H_1 \otimes ... \otimes H_p) = T\left(E(H_1 \otimes ... \otimes H_p)\right), \tag{6.25}$$

a formula, which will be very useful to us.

8. Moreover, if the bt A is symmetrical then we can easily see the for every tensor T,

$$T(A) = T^{(S)}(A). \tag{6.26}$$

In particular, if we have an arbitrary random bt H (with a sufficient number of moments) then $E(H \otimes ... \otimes H)$ is symmetrical, hence (6.26) implies that, for all bt T, $T(E(H \otimes ... \otimes H)) = T^{(S)}(E(H \otimes ... \otimes H))$.

9. As we already stated, the bts form a finite dimensional vector space. Hence all norms are equivalent, in the sense that the ratio between two norms is (for all elements of the reference space with the exception of 0) bounded from above and bounded from below with a bound strictly bigger than zero. Hence convergence properties of sequences are the same for different norms, and we do not need to care which norm we use. Of particular interest, however, is the norm

$$\|T\| = \sqrt{\sum t^2_{i_1,...,i_p}}, \tag{6.27}$$

where the $t_{i_1,...,i_p}$ are the coordinates of T. Cauchy-Schwartz inequality and (6.23) imply that for all bts T, C :

$$|T(C)| \leq \|T\| \, \|C\|. \tag{6.28}$$

Estimates for the norms of tensor products are more difficult - we will discuss them later on when they appear.

6.2. Other notations

Definition. $\mathcal{H}_{t,T}$ is defined as the σ-algebra generated by $(\eta_t, \eta_{t-1}, ...\eta_1, y_T, .., y_1)$.Then $\mathcal{H}_{0,T}$ is the σ-algebra generated by the data only.

The sample is split in the following way:

$$t = \underbrace{1, 2, ..., T_1}_{\text{1st block}}, \underbrace{T_1 + 1, ..., T_2}_{\text{2d block}}, ..., \underbrace{T_{i-1} + 1, ..., T_i}_{i\text{th block}}, ..., \underbrace{T_{B_N - 1} + 1, ..., T_{B_N}}_{B_N\text{th block}}$$

There are B_N blocks and each block has B_L or $B_L - 1$ elements. i is the index for the block $i = 1, ..., B_N$. We use the convention $T_0 = 0$ and $T_{B_N} = T$. In the sequel we will decompose the sum as follows:

$$\sum_{t=1}^{T} = \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} .$$

Our analysis is based on the derivatives of the logarithm of the likelihood function. We did denote the conditional parametric densities by $f_t = f_t(\theta_T)$, and the log-likelihood functions by l_t. We did define

$$D^k l_t = l_t^{(k)} \tag{6.29}$$

First we need to derive the tensorized forms of well-known Bartlett identities (Bartlett, 1953a,b). Let us define for an arbitrary, but fixed h the function

$$\ell_t(u) = \log f_t\left(\theta_T + uh\right) \tag{6.30}$$

When differentiating ℓ_t , one obtains - with $f = .f_t(\theta_T)$ and $f', f^{(2)}, ..$ denoting the derivatives of $f_t\left(\theta_T + uh\right)$ with respect to u.

1st derivative: $\ell_t^{(1)} = \dfrac{f'}{f}$.

2nd derivative: $\ell_t^{(2)} = \dfrac{f^{(2)}}{f} - \dfrac{f'}{f^2} f'$.

3rd derivative: $\ell_t^{(3)} = \dfrac{f^{(3)}}{f} - \dfrac{f^{(2)}}{f^2} f' - \dfrac{2f' f^{(2)}}{f^2} + 2\dfrac{f'^2}{f^3} f'$.

4th derivative: $\ell_t^{(4)} = \dfrac{f^{(4)}}{f} - \dfrac{f^{(3)}}{f^2} f' - \dfrac{3f^{(3)} f' + 3f^{(2)} f^{(2)}}{f^2} + 6\dfrac{f^{(2)} f'}{f^3} f' + 6\dfrac{f'^2}{f^3} f^{(2)} - 6\dfrac{f'^3}{f^4} f'$

According to the formalism outlined previously, we can conclude that $\ell_t^{(k)} = l_t^{(k)}(h, ..h)$ and that $f^{(k)} = D^k f(h, ..h)$. Taking into account our characterization of the tensor product (6.22), and the technique described under topic 4 in the list of properties of bts, we can conclude that

$$l_t^{(1)} = (1/f_t) D f_t \tag{6.31}$$

$$\frac{1}{f_t} D^2 f_t = l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)}, \tag{6.32}$$

$$\frac{1}{f_t} D^3 f_t = \left(l_t^{(3)} + 3 l_t^{(2)} \otimes l_t^{(1)} + l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \right)^{(S)} \tag{6.33}$$

$$\frac{1}{f_t} D^4 f_t = \left(l_t^{(4)} + 6 l_t^{(2)} \otimes l_t^{(1)} \otimes l_t^{(1)} + 4 l_t^{(3)} \otimes l_t^{(1)} + 3 l_t^{(2)} \otimes l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \right)^{(S)}. \tag{6.34}$$

We can easily see that we do not need to symmetrize (6.32), since the bt on the right hand side is symmetrical. Let us now denote by $\mathcal{F}_t$ the σ-algebra generated by the data $y_t, y_{t-1}, ...$Then one can easily see that for $k \leq 4$, we have for arbitrary h $E(\frac{1}{f_t} D^k f_t\,(h, ..h)) \,/ \mathcal{F}_{t-1}) = \int \frac{1}{f_t} D^k f_t(h, ..h) f_t d\mu(y_t) = \int D^k f_t(h, ..h) d\mu(y_t)$, where μ is the dominating measure defined in Section 2. Since we assumed f_t to be at least 5 times differentiable (and the 5th derivative to be uniformly integrable), we can easily see (cf...) that we can interchange integral and differentiation, and conclude that $\int D^k f_t(h, ..., h) d\mu(y_t) = D^k (\int f_t d\mu(y_t))(h, ..., h) = 0$, since all the f_t as conditional densities integrate to one.

Let us define

$$\begin{aligned}
m_{2,t} &= \left(l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)} + 2 l_t^{(1)} \otimes L_{t-1} \right)^{(S)} \\
m_{3,t} &= \left(l_t^{(3)} + 3 l_t^{(2)} \otimes l_t^{(1)} + l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \right)^{(S)} \\
m_{4,t} &= \left(l_t^{(4)} + 6 l_t^{(2)} \otimes l_t^{(1)} \otimes l_t^{(1)} + 4 l_t^{(3)} \otimes l_t^{(1)} + 3 l_t^{(2)} \otimes l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \right)^{(S)}
\end{aligned}$$

where $L_t = \sum_{s=T_{i-1}+1}^{t} l_t^{(1)}$, and furthermore

$$M_j = \sum_{t=T_{i-1}+1}^{T_i} m_{j,t}, j = 2, 3, 4. \tag{6.35}$$

Then we can easily see that $l_t^{(1)}$, $m_{2,t}$, $m_{3,t}$, $m_{4,t}$ are martingale difference sequences with respect to the $\mathcal{F}_t$. Furthermore, the $m_{j,t}$,$j = 2, 3, 4$ are defined as symmetrizations of the bts on the rhs of the above equations.

In the sequel, we will heavily rely on (6.26), both for the evaluation of the $m_{j,t}$,$j = 2, 3, 4$ and the derivatives as well: When we evaluate bts with symmetrical arguments, it is irrelevant whether we use the bts themselves or the nonsymmetrical expressions used in the above definitions.

7. Appendix B: Proofs

The first theorem we want to prove is 3.1. The statement of the theorem involves some uniform convergence in probability of a parametrized family of random variables. So let us assume the theorem would not be true. There existed a compact subset $K \subseteq \Theta \times B$ so that we do not have uniform convergence in probability on K. Then there exists a sequence $(\theta_T, \beta_T) \in K$ and a $\varepsilon > 0$ so that

$$P_{\theta_T}\left(\left[\left|\ell_T^{\beta_T}(\theta_T) / \exp\left(\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\theta_T,\beta_T) - \frac{1}{8}E\left(\mu_{2,t}(\theta_T,\beta_T)^2\right)\right) - 1\right| \geq \varepsilon\right]\right) \geq \varepsilon. \tag{7.1}$$

Since the (θ_T, β_T) are elements of a compact subset, we there is a convergent subsequence. Hence **to prove** theorem 3.1, it is sufficient to show that for every $(\theta_T, \beta_T) \to (\theta_0, \beta_0)$

$$P_{\theta_T}\left(\left[\left|\ell_T^{\beta_T}(\theta_T) / \exp\left(\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\theta_T,\beta_T) - \frac{1}{8}E\left(\mu_{2,t}(\theta_T,\beta_T)^2\right)\right) - 1\right| \geq \varepsilon\right]\right) \to 0. \tag{7.2}$$

or

$$\ell_T^{\beta_T}(\theta_T) / \exp\left(\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\theta_T,\beta_T) - \frac{1}{8}E\left(\mu_{2,t}(\theta_T,\beta_T)^2\right)\right) \to 1 \text{ in probability with respect to } P_{\theta_T}. \tag{7.3}$$

In the sequel, we will prove this relationship. To simplify our notation, however, we will suppress the parameters (θ_T, β_T) and (θ_0, β_0). When analyzing expressions related to a sample of length T, we simply write E and P instead of E_{θ_T} and P_{θ_T}. Moreover, we also will drop the argument from expression like $l_t(\theta_T)$, ..and simply use l_t, ... The proper argument should be evident from the context. This simplification of notations brings significant advantages for our calculations of derivatives: When we are using arguments in connection with derivatives then they are meant to be arguments of the corresponding multilinear form. As an example, the expression $l_t^{(2)}$should denote the second derivative of l_t at θ_T, which is a bilinear form. So $l_t^{(2)}(h,k)$ is the evaluation of this bilinear form with the arguments h and k.

The following lemmas are used in the proof of Theorem 3.1

Lemma 7.1. *For any $\varepsilon > 0$, we can find $1-\varepsilon \leq \frac{f_T}{f_T^*} \leq 1+\varepsilon$ on some set A_T^ε so that $\lim_{T\to\infty}\sup P(A_T^\varepsilon) = 1$ where A_T^ε is $\mathcal{H}_{0,T}$-measurable and independent on β. Then* $\dfrac{\sup_\beta E(f_T|\mathcal{H}_{0,T})}{\sup_\beta E(f_T^*|\mathcal{H}_{0,T})} \xrightarrow{P} 1.$

A sufficient condition for Lemma 7.1 is

$$\left|\frac{f_T}{f_T^*}\right| \leq 1 + B_T$$

where B_T is $\mathcal{H}_{0,T}$-measurable and independent on β and $B_T \xrightarrow{P} 0$.

Lemma 7.2. *Let x_i be$\mathcal{H}_{T_i,T}$ measurable random variables and let $\Delta_{i,T} = E\left(x_i | \mathcal{H}_{T_{i-1},T}\right)$. Assume there are bounds B_T and $C_T \to 0$ $\mathcal{H}_{0,T}-$measurable and independent of β such that*

$$\sup_{\beta} \left| \sum_{i=1}^{B_N} \Delta_{i,T} \right| \leq B_T \tag{7.4}$$

and

$$\sup_{\beta} \sum_{i=1}^{B_N} \Delta_{i,T}^2 \leq C_T. \tag{7.5}$$

Then

$$\sup_{\beta} E\left[\prod_{i=1}^{B_N} (1 + x_i) | \mathcal{H}_{0,T}\right] \xrightarrow{P} 1. \tag{7.6}$$

Lemma 7.3. *Let $\Delta_{i,T} = E\left(x_i | \mathcal{H}_{T_{i-1},T}\right)$. Assume there is an $\mathcal{H}_{0,T}-$measurable set A_T so that $\|x_i\| \leq 1/2$ on A_T and $P(A_T) \to 1$. Moreover, assume that $\Delta_{i,T}$ is a martingale with respect to the data and*

$$\sup_{\beta} \sum_{i=1}^{B_N} E\Delta_{i,T}^2 \to 0.$$

Then (7.6) is satisfied.

Lemma 7.4. *Let $a_1, a_2, ..., a_N$ a sequence of numbers with $N = 1, 2, ...$ Then*

$$\left(\sum_{i=1}^{N} |a_i|\right)^l \leq N^{l-1} \sum_{i=1}^{N} |a_i|^l, \; l = 1, 2, ...$$

Lemma 7.5. *A sufficient condition for Conditions (7.4) and (7.5) is*

$$\sum_i E |\Delta_i| \to 0. \tag{7.7}$$

The following lemma gives a result for the product of 4 arbitrary terms x_{ij}. The index is denoted $j = 1, 2, 3, 4$ for convenience.

Lemma 7.6. *Let $x_{ij} = \sum_t \widetilde{x}_{ijt} / T^{\alpha_j}$. Assume that*

$$E\left(\left|\sum_t \widetilde{x}_{ijt}\right|^4\right) \leq B_L^m \tag{7.8}$$

for some $m \geq 1$ *and all* $j = 1, 2, 3, 4$. *Let* $k \leq 4$ *and* $\mathcal{D}=\{d_1, ..., d_k\}$ *be any* $k-$*partition of the integers* $1, 2, 3, 4$. *Assume that*

$$\sum_{j \in \mathcal{D}} \alpha_j > 1, \tag{7.9}$$

and let B_L *be such that*

$$\sum_i E\left(\prod_{j \in \mathcal{D}} |x_{ij}|\right) = o(1).$$

Then, Conditions (7.4) and (7.5) are satisfied for $\Delta_i = E\left(\prod_{j\in\mathcal{D}} x_{ij} | \mathcal{H}_{T_{i-1},T}\right)$.

Lemma 7.7. *Assume Assumption 4 holds. Let const denotes a constant independent of* β*, we have*

$$\begin{aligned}
E\left(\|\tilde{x}_{i1}\|^4\right) &= E\left(\|L_1\|^4\right) \leq const B_L^4 \\
E\left(\|\tilde{x}_{i2}\|^4\right) &= E\left(\|M_2\|^4\right) \leq const B_L^8, \\
E\left(\|\tilde{x}_{i3}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)} m_{2,t}\right\|^4\right) \leq const B_L^8, \\
E\left(\|\tilde{x}_{i4}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)2} m_{2,t}\right\|^4\right) \leq const B_L^8, \\
E\left(\|\tilde{x}_{i5}\|^4\right) &= E\left(\|M_3\|^4\right) \leq const B_L^4, \\
E\left(\|\tilde{x}_{i6}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)} m_{3,t}\right\|^4\right) \leq const B_L^4, \\
E\left(\|\tilde{x}_{i7}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)} L_{t-1}^2\right\|^4\right) \leq const B_L^{12}, \\
E\left(\|\tilde{x}_{i8}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)2} L_{t-1}^2\right\|^4\right) \leq const B_L^{12}, \\
E\left(\|\tilde{x}_{i9}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)} L_{t-1}^3\right\|^4\right) \leq const B_L^{16}, \\
E\left(\|\tilde{x}_{i10}\|^4\right) &= E\left(\left\|\sum_t l_t^{(1)} L_{t-1} m_{2,t}\right\|^4\right) \leq const B_L^{16},
\end{aligned}$$

Proof of Lemma 7.1. Let η be an arbitrary positive number and $0 < \varepsilon < \eta$.

$$\begin{aligned}\sup_\beta E(f_T|\mathcal{H}_{0,T}) &= \sup_\beta E(\frac{f_T}{f_T^*} f_T^*|\mathcal{H}_{0,T}) \\ &= I_{A_T^\varepsilon} \sup_\beta E(\frac{f_T}{f_T^*} f_T^*|\mathcal{H}_{0,T}) + I_{\left(A_T^\varepsilon\right)^c} \sup_\beta E(\frac{f_T}{f_T^*} f_T^*|\mathcal{H}_{0,T}) .\end{aligned}$$

Under of the assumptions of the lemma:

$$\begin{aligned}& I_{A_T^\varepsilon} \ (1-\varepsilon) \sup_\beta E(f_T^*|\mathcal{H}_{0,T}) + I_{\left(A_T^\varepsilon\right)^c} \sup_\beta E(f_T|\mathcal{H}_{0,T}) \\ \leq\ & \sup_\beta E(f_T|\mathcal{H}_{0,T}) \\ \leq\ & I_{A_T^\varepsilon} \ (1+\varepsilon) \sup_\beta E(f_T^*|\mathcal{H}_{0,T}) + I_{\left(A_T^\varepsilon\right)^c} \sup_\beta E(f_T|\mathcal{H}_{0,T}).\end{aligned}$$

To simplify notation, we denote $\dfrac{\sup_\beta E(f_T|\mathcal{H}_{0,T})}{\sup_\beta E(f_T^*|\mathcal{H}_{0,T})}$ by X_T, then

$$I_{A_T^\varepsilon} \ (1-\varepsilon) + I_{\left(A_T^\varepsilon\right)^c} X_T \leq X_T \leq I_{A_T^\varepsilon} \ (1+\varepsilon) + I_{\left(A_T^\varepsilon\right)^c} X_T$$

$$\begin{aligned}& P\left[|X_T - 1| < \eta\right] \\ =\ & P\left[1-\eta < X_T < 1+\eta\right] \\ \geq\ & P\left[\left\{I_{A_T^\varepsilon} (1+\varepsilon) + I_{\left(A_T^\varepsilon\right)^c} X_T < 1+\eta\right\} \cap \left\{I_{A_T^\varepsilon} (1-\varepsilon) + I_{\left(A_T^\varepsilon\right)^c} X_T > 1-\eta\right\}\right] \\ =\ & P\left(A_T^\varepsilon\right) + P\left(\left(A_T^\varepsilon\right)^c\right) P\left[1-\eta < X_T < 1+\eta\right] \\ \geq\ & P\left(A_T^\varepsilon\right) \to 1.\end{aligned}$$

Hence $X_T \xrightarrow{P} 1$.

Proof of Lemma 7.2. Using a Taylor expansion, we see that Conditions (7.4) and (7.5) imply that

$$\sum_{i=1}^{B_N} \ln\left(1+\Delta_{i,T}\right) = \sum_{i=1}^{B_N} \Delta_{i,T} - \frac{\Delta_{i,T}^2}{2} + o\left(\Delta_{i,T}^2\right) \xrightarrow{P} 0$$

or more precisely

$$1-\varepsilon \leq \prod_{i=1}^{B_N} \left(1+\Delta_{i,T}\right) \leq 1+\varepsilon$$

for any $\varepsilon > 0$ on a set A_T^ε $\mathcal{H}_{0,T}$-measurable and independent of β such that $P\left(A_T^\varepsilon\right) \to 1$.

Using iterated expectations, we obtain

$$E\left[\frac{\prod_{i=1}^{B_N}\left(1+x_i\right)}{\prod_{i=1}^{B_N}\left(1+\Delta_{i,T}\right)} |\mathcal{H}_{0,T}\right] = 1$$

Hence on A_T^ε, we have

$$\frac{1}{1+\varepsilon}\sup_\beta E\left[\prod_{i=1}^{B_N}(1+x_i)\,|\mathcal{H}_{0,T}\right] \leq 1 \leq \frac{1}{1-\varepsilon}\sup_\beta E\left[\prod_{i=1}^{B_N}(1+x_i)\,|\mathcal{H}_{0,T}\right]$$

or equivalently

$$1-\varepsilon \leq \sup_\beta E\left[\prod_{i=1}^{B_N}(1+x_i)\,|\mathcal{H}_{0,T}\right] \leq 1+\varepsilon.$$

As $P\left(A_T^\varepsilon\right) \to 1$, it follows that $\sup_\beta E\left[\prod_{i=1}^{B_N}(1+x_i)\,|\mathcal{H}_{0,T}\right] \xrightarrow{P} 1.$

Proof of Lemma 7.3.

$$\begin{aligned}
&\ln E\left[\prod_{i=1}^{B_N}(1+x_i)\,|\mathcal{H}_{0,T}\right] \\
=& \sum_{l=1}^{B_N}\left\{\ln E\left[\prod_{i=1}^{l}(1+x_i)\,|\mathcal{H}_{0,T}\right] - \ln E\left[\prod_{i=1}^{l-1}(1+x_i)\,|\mathcal{H}_{0,T}\right]\right\} \\
=& \sum_{l=1}^{B_N}\left\{\ln\left(u_l+h_l\right)-\ln\left(u_l\right)\right\}
\end{aligned}$$

where

$$\begin{aligned}
u_l &= E\left[\prod_{i=1}^{l-1}(1+x_i)\,|\mathcal{H}_{0,T}\right], \\
h_l &= E\left[\prod_{i=1}^{l}(1+x_i)\,|\mathcal{H}_{0,T}\right] - E\left[\prod_{i=1}^{l-1}(1+x_i)\,|\mathcal{H}_{0,T}\right] \\
&= E\left[x_l\prod_{i=1}^{l-1}(1+x_i)\,|\mathcal{H}_{0,T}\right] \\
&= E\left[E\left(x_l|\mathcal{H}_{T_{l-1},T}\right)\prod_{i=1}^{l-1}(1+x_i)\,|\mathcal{H}_{0,T}\right] \\
&= E\left[\Delta_l\prod_{i=1}^{l-1}(1+x_i)\,|\mathcal{H}_{0,T}\right].
\end{aligned}$$

Using a Taylor expansion, we have

$$
\begin{aligned}
\left|\sum_{l=1}^{B_N}\left\{\ln\left(u_l+h_l\right)-\ln\left(u_l\right)-\frac{h_l}{u_l}\right\}\right| &\leq \sum_{l=1}^{B_N}\frac{h_l^2}{2\left(\left|u_l\right|^2-\left|h_l\right|^2\right)} \\
&= \frac{1}{2}\sum_{l=1}^{B_N}\left(\frac{h_l}{u_l}\right)^2\frac{1}{1-\left(\frac{h_l}{u_l}\right)^2} \\
&\leq \frac{1}{2}\sum_{l=1}^{B_N}\left(\frac{h_l}{u_l}\right)^2\frac{1}{1-\sum_{l=1}^{B_N}\left(\frac{h_l}{u_l}\right)^2}. \qquad (7.10)
\end{aligned}
$$

Let $\delta_l = h_l/u_l$. If we are able to show that

$$\sum_{l=1}^{B_N}\delta_l \xrightarrow{P} 0, \qquad (7.11)$$

$$\sum_{l=1}^{B_N}\delta_l^2 \xrightarrow{P} 0, \qquad (7.12)$$

then we have

$$\left|\sum_{l=1}^{B_N}\left\{\ln\left(u_l+h_l\right)-\ln\left(u_l\right)\right\}\right| \xrightarrow{P} 0,$$

which itself implies

$$E\left[\prod_{i=1}^{B_N}\left(1+x_i\right)|\mathcal{H}_{0,T}\right] \xrightarrow{P} 1.$$

(7.11) will follow from (7.12) and the fact that δ_l is a martingale as Δ_l is itself a martingale. Now we want to show that

$$\sum_{l=1}^{B_N}E\left(\Delta_l^2\right)\to 0 \Rightarrow \sum_{l=1}^{B_N}E\left(\delta_l^2\right)\to 0 \Rightarrow \sum_{l=1}^{B_N}\delta_l^2 \xrightarrow{P} 0. \qquad (7.13)$$

We have

$$
\begin{aligned}
\delta_l &= \frac{E\left[\Delta_l\prod_{i=1}^{l-1}\left(1+x_i\right)|\mathcal{H}_{0,T}\right]}{E\left[\prod_{i=1}^{l-1}\left(1+x_i\right)|\mathcal{H}_{0,T}\right]} \\
&= \frac{E\left[\Delta_l\prod_{i=1}^{l-2}\left(1+x_i\right)\left(1+x_{l-1}\right)|\mathcal{H}_{0,T}\right]}{E\left[\prod_{i=1}^{l-2}\left(1+x_i\right)\left(1+x_{l-1}\right)|\mathcal{H}_{0,T}\right]} \\
&\leq \frac{3E\left[\Delta_l\prod_{i=1}^{l-2}\left(1+x_i\right)|\mathcal{H}_{0,T}\right]}{E\left[\prod_{i=1}^{l-2}\left(1+x_i\right)|\mathcal{H}_{0,T}\right]}
\end{aligned}
$$

because $\|x_{l-1}\| \leq 1/2$ by assumption. Note that $E\left(\Delta_l|\mathcal{H}_{T_{l-2},T}\right) \leq E\left(|\Delta_l|\,|\mathcal{H}_{T_{l-2},T}\right)$ and it follows from the geometric ergodicity of η_t that

$$\left|E\left(|\Delta_l|\,|\mathcal{H}_{T_{l-2},T}\right) - E\left(|\Delta_l|\,|\mathcal{H}_{0,T}\right)\right| \leq \lambda^{B_L} g\left(\mathcal{H}_{T_{l-2},T}\right)$$

where g is some positive integrable function of $\mathcal{H}_{T_{l-2},T}$. Hence

$$\begin{aligned} E\left(|\Delta_l|\,|\mathcal{H}_{T_{l-2},T}\right) &\leq \left|E\left(|\Delta_l|\,|\mathcal{H}_{T_{l-2},T}\right) - E\left(|\Delta_l|\,|\mathcal{H}_{0,T}\right)\right| + \left|E\left(|\Delta_l|\,|\mathcal{H}_{0,T}\right)\right| \\ &\leq \lambda^{B_L} g\left(\mathcal{H}_{T_{l-2},T}\right) + \left|E\left(|\Delta_l|\,|\mathcal{H}_{0,T}\right)\right|. \end{aligned}$$

$$\begin{aligned} \delta_l &\leq \frac{3E\left[E\left(|\Delta_l|\,|\mathcal{H}_{T_{l-2},T}\right)\prod_{i=1}^{l-2}(1+x_i)\,|\mathcal{H}_{0,T}\right]}{E\left[\prod_{i=1}^{l-2}(1+x_i)\,|\mathcal{H}_{0,T}\right]} \\ &\leq 3\lambda^{B_L}\frac{E\left[g\left(\mathcal{H}_{T_{l-2},T}\right)\prod_{i=1}^{l-2}(1+x_i)\,|\mathcal{H}_{0,T}\right]}{E\left[\prod_{i=1}^{l-2}(1+x_i)\,|\mathcal{H}_{0,T}\right]} + 3E\left(|\Delta_l|\,|\mathcal{H}_{0,T}\right), \\ \delta_l^2 &\leq O\left(\lambda^{B_L}\right) + 9E\left(|\Delta_l|^2\,|\mathcal{H}_{0,T}\right). \end{aligned}$$

We get

$$\sum_{l=1}^{B_N} E\left(\delta_l^2\right) \leq O\left(B_N\lambda^{B_L}\right) + 9\sum_{l=1}^{B_N} E\left(|\Delta_l|^2\right).$$

This proves the first implication of (7.13). The second implication follows from Markov's inequality.

Proof of Lemma 7.4. Let $p_i = |a_i| / \sum_{i=1}^N |a_i|$. The problem consists in solving

$$\min_{p_i} \sum_{i=1}^N p_i^l$$

subject to $\sum_{i=1}^N p_i = 1$. The solution is $\sum_{i=1}^N p_i^l = 1/N^{l-1}$.

Proof of Lemma 7.5 (a) (7.7) implies $\sum_i |\Delta_i| \overset{P}{\to} 0$ by Markov's theorem, hence as $|\sum_i \Delta_i| \leq \sum_i |\Delta_i|$, Condition (7.4) follows, (b) $\sum_i |\Delta_i| \overset{P}{\to} 0$ means that for T large enough, $|\Delta_i| < 1$, and hence $|\Delta_i|^2 \leq |\Delta_i|$, therefore Condition (7.5) follows.

Proof of Lemma 7.6. By the geometric-arithmetic mean inequality, we have

$$\begin{aligned} E\left(\prod_{j=1}^k |x_{ij}|\right) &= E\left(\sqrt[k]{|x_{i1}|^k \ldots |x_{ik}|^k}\right) \leq \frac{1}{k}\sum_{j=1}^k E\left(|x_{ij}|^k\right) \\ &\leq \frac{B_L^m}{T^{\sum_{j=1}^k \alpha_j}}. \end{aligned}$$

Hence

$$\sum_i E\left(\prod_{j=1}^k |x_{ij}|\right) \leq \frac{T}{B_L}\frac{B_L^m}{T^{\sum_{j=1}^k \alpha_j}} = \frac{B_L^{m-1}}{T^{\sum_{j=1}^k \alpha_j - 1}} = o(1).$$

The last statement of the lemma follows from $E\left|\Delta_i\right| \leq E\left[\left|E\left(\prod_{j=1}^k x_{ij}|\mathcal{H}_{T_{i-1},T}\right)\right|\right] \leq E\left[E\left(\left|\prod_{j=1}^k x_{ij}\right| |\mathcal{H}_{T_{i-1},T}\right)\right] = E\left(\left|\prod_{j=1}^k x_{ij}\right|\right)$.

Proof of Lemma 7.7. Term $\widetilde{x}_{i1}$

$$E\left|L_1\right|^4 \leq B_L^3 \sum_{t=T_i+1}^{T_{i+1}} E\left\|l_t^{(1)4}\right\| \leq B_L^4 \sup_t E\left\|l_t^{(1)}\right\|^4$$

by Lemma 7.4.

Term $\tilde{x}_{i2}$:

$$\begin{aligned}
E\left(\left|M_2\right|^4\right) &\leq B_L^3 \sum_t E\left(\left|m_{2,t}\right|^4\right) \\
&\leq B_L^3 \sum_t E\left(\left\|l_t^{(2)}\right\|^4 + \left\|l_t^{(1)}\right\|^8 + 2^4\left\|l_t^{(1)}\right\|^4 \left\|L_{t-1}\right\|^4\right) \\
&\leq B_L^3 \sum_t \left(E\left\|l_t^{(2)}\right\|^4 + E\left\|l_t^{(1)}\right\|^8 + 2^4\left(E\left(\left\|l_t^{(1)}\right\|^8\right) E\left(\left\|L_{t-1}\right\|^8\right)\right)^{1/2}\right) \\
&\leq const B_L^8
\end{aligned}$$

provided $\sup E\left\|l_t^{(1)}\right\|^8 < \infty$ and $\sup E\left\|l_t^{(2)}\right\|^4 < \infty$.

Term $\widetilde{x}_{i3}$:

$$\begin{aligned}
& E\left(\left|\sum_t l_t^{(1)} m_{2,t}\right|^4\right) \\
=\; & E\left(\left|\sum_t l_t^{(1)} l_t^{(2)} + l_t^{(1)3} + 2 l_t^{(1)2} L_{t-1}\right|^4\right) \\
\leq\; & B_L^4 \sum_t E\left(\left\|l_t^{(1)} l_t^{(2)}\right\|^4 + \left\|l_t^{(1)}\right\|^{12} + 2^4\left\|l_t^{(1)2} L_{t-1}\right\|^4\right) \\
\leq\; & B_L^4 \sum_t \left(E\left\|l_t^{(1)}\right\|^8 E\left\|l_t^{(2)}\right\|^8\right)^{1/2} + E\left\|l_t^{(1)}\right\|^{12} + 2^4\left(E\left\|l_t^{(1)}\right\|^8 E\left(\left\|L_{t-1}\right\|^8\right)\right)^{1/2} \\
\leq\; & const B_L^8
\end{aligned}$$

provided that $\sup E\left\|l_t^{(1)}\right\|^{12} < \infty$ and $\sup E\left\|l_t^{(2)}\right\|^8 < \infty$.

Term $\widetilde{x}_{i4}$:

$$\begin{aligned}
& E\left(\left|\sum_t l_t^{(1)2} m_{2,t}\right|^4\right) \\
= \; & E\left(\left|\sum_t l_t^{(1)2} l_t^{(2)} + l_t^{(1)4} + 2 l_t^{(1)3} L_{t-1}\right|^4\right) \\
\leq \; & B_L^3 \sum_t E\left(\left\| l_t^{(1)2} l_t^{(2)}\right\|^4 + \left\| l_t^{(1)}\right\|^{16} + 2^4 \left\| l_t^{(1)3} L_{t-1}\right\|^4\right) \\
\leq \; & B_L^3 \sum_t \left(E\left\| l_t^{(1)}\right\|^{16} E\left\| l_t^{(2)}\right\|^8\right)^{1/2} + E\left\| l_t^{(1)}\right\|^{16} + 2^4 \left(E\left\| l_t^{(1)}\right\|^{24} E\left(\left\| L_{t-1}\right\|^8\right)\right)^{1/2} \\
\leq \; & const B_L^8
\end{aligned}$$

provided that $\sup E\left\| l_t^{(1)}\right\|^{24} < \infty$ and $\sup E\left\| l_t^{(2)}\right\|^8 < \infty$.

Term $\widetilde{x}_{i5}$:

$$\begin{aligned}
E\left|M_3\right|^4 \; & \leq \; B_L^3 \sum_t E\left(\left\| l_t^{(3)}\right\|^4 + \left\| l_t^{(2)} l_t^{(1)}\right\|^4 + \left\| l_t^{(1)3}\right\|^4\right) \\
& \leq \; B_L^3 \sum_t E\left\| l_t^{(3)}\right\|^4 + \left(E\left\| l_t^{(2)}\right\|^8 E\left\| l_t^{(1)}\right\|^8\right)^{1/2} + E\left\| l_t^{(1)}\right\|^{12} \\
& \leq \; const B_L^4
\end{aligned}$$

provided $\sup E\left\| l_t^{(1)}\right\|^{12} < \infty$, $\sup E\left\| l_t^{(2)}\right\|^8 < \infty$, and $\sup E\left\| l_t^{(3)}\right\|^4 < \infty$.

Term $\widetilde{x}_{i6}$:

$$\begin{aligned}
E\left(\left|\sum_{t=T_i+1}^{T_{i+1}} m_{3,t} l_t^{(1)}\right|^4\right) \; & \leq \; B_L^4 \sum_t E\left| m_{3,t} l_t^{(1)}\right|^4 \\
& \leq \; B_L^4 \sum_t E\left(\left\| l_t^{(3)} l_t^{(1)}\right\|^4 + \left\| l_t^{(2)} l_t^{(1)2}\right\|^4 + \left\| l_t^{(1)4}\right\|^4\right) \\
& \leq \; const B_L^4
\end{aligned}$$

provided that $\sup E\left\| l_t^{(1)}\right\|^{16} < \infty$, $\sup E\left\| l_t^{(2)}\right\|^8 < \infty$ and $\sup E\left\| l_t^{(3)}\right\|^8 < \infty$.

Term $\widetilde{x}_{i7}$:

$$\begin{aligned} E\left(\left|\sum_t l_t^{(1)} L_{t-1}^2\right|^4\right) &\leq B_L^3 \sum_t E\left|l_t^{(1)} L_{t-1}^2\right|^4 \\ &\leq B_L^3 \sum_t \left(E\left\|l_t^{(1)}\right\|^8 E\left\|L_{t-1}\right\|^{16}\right)^{1/2} \\ &\leq const B_L^{12} \end{aligned}$$

provided $\sup E\left\|l_t^{(1)}\right\|^{16} < \infty$.

Term $\widetilde{x}_{i8}$:

$$\begin{aligned} E\left(\left|\sum_t l_t^{(1)2} L_{t-1}^2\right|^4\right) &\leq B_L^3 \sum_t E\left|l_t^{(1)2} L_{t-1}^2\right|^4 \\ &\leq B_L^3 \sum_t \left(E\left\|l_t^{(1)}\right\|^{16} E\left\|L_{t-1}\right\|^{16}\right)^{1/2} \\ &\leq const B_L^{12} \end{aligned}$$

provided $\sup E\left\|l_t^{(1)}\right\|^{16} < \infty$.

Term $\widetilde{x}_{i9}$:

$$\begin{aligned} E\left(\left|\sum_t l_t^{(1)} L_{t-1}^3\right|^4\right) &\leq B_L^3 \sum_t E\left|l_t^{(1)} L_{t-1}^3\right|^4 \\ &\leq B_L^3 \sum_t \left(E\left\|l_t^{(1)}\right\|^8 E\left\|L_{t-1}\right\|^{24}\right)^{1/2} \\ &\leq const B_L^{16} \end{aligned}$$

provided $\sup E\left\|l_t^{(1)}\right\|^{24} < \infty$.

Term $\widetilde{x}_{i10}$:

$$
\begin{aligned}
E\left(\left|\sum_t l_t^{(1)} L_{t-1} m_{2,t}\right|^4\right) &= E\left|\sum_t l_t^{(1)} l_t^{(2)} L_{t-1} + l_t^{(1)3} L_{t-1} + 2 l_t^{(1)2} L_{t-1}^2\right|^4 \\
&\leq B_L^3 \sum_t E\left|l_t^{(1)} l_t^{(2)} L_{t-1}\right|^4 + E\left|l_t^{(1)3} L_{t-1}\right|^4 + 2^4 E\left|l_t^{(1)2} L_{t-1}^2\right|^4 \\
&\leq B_L^3 \sum_t \left\{ E\left\|l_t^{(1)}\right\|^{12} + E\left\|l_t^{(2)}\right\|^{12} + E\left\|L_{t-1}\right\|^{12} \right. \\
&\quad \left. + \left(E\left\|l_t^{(1)}\right\|^{24} E\left\|L_{t-1}\right\|^8\right)^{1/2} + 2^4\left(E\left\|l_t^{(1)}\right\|^{16} E\left\|L_{t-1}\right\|^{16}\right)^{1/2} \right\} \\
&\leq const B_L^{16}
\end{aligned}
$$

provided $\sup E\left\|l_t^{(1)}\right\|^{24} < \infty$ and $\sup E\left\|l_t^{(2)}\right\|^{12} < \infty$.

Proof of Theorem 3.1

Denote TE_T the Taylor expansion of $\sum_t \left(l_t\left(\theta_T + \eta_t/T^{1/4}\right) - l_t\left(\theta_T\right)\right)$ around θ_T :

$$
TE_T = \sum_{t=1}^{T}\left[\frac{1}{\sqrt[4]{T}} l_t^{(1)}(\eta_t) + \frac{1}{2\sqrt{T}} l_t^{(2)}(\eta_t, \eta_t) + \frac{1}{6\sqrt[4]{T^3}} l_t^{(3)}(\eta_t, \eta_t, \eta_t) + \frac{1}{24T} l_t^{(4)}(\eta_t, \eta_t, \eta_t, \eta_t)\right] \tag{7.14}
$$

where $l_t^{(1)}, ..., l_t^{(4)}$ are function of θ_T. The proof is in three steps.

Denote

$$
\widetilde{TS}_T\left(\beta, \theta\right) = \frac{1}{2}\frac{1}{\sqrt{T}}\sum_t \mu_{2,t}\left(\beta, \theta\right) - \frac{1}{8}\frac{1}{T}\sum_t \left[\mu_{2,t}\left(\theta, \beta\right)\right]^2.
$$

Step 1. Using Lemma 7.1, we show that

$$
\frac{\ell_T\left(\theta, \beta\right)}{E\left[\exp\left(TE_T\right)|\mathcal{H}_{0,T}\right]} \xrightarrow{P} 1
$$

uniformly in β.

Step 2. Using Lemma 7.1, we show that

$$
\frac{E\left[\exp\left(TE_T\right)|\mathcal{H}_{0,T}\right]}{E\left[\exp\left(\widetilde{TS}_T + \sum_{i=1}^{B_N}\sum_{j=1}^{J}\ln\left(1+x_{ij}\right)\right)|\mathcal{H}_{0,T}\right]} \xrightarrow{P} 1
$$

uniformly in β for some x_{ij}.

Step 3. Using Lemma 7.2, we prove that

$$
\frac{E\left[\exp\left(\widetilde{TS}_T + \sum_{i=1}^{B_N}\sum_{j=1}^{J}\ln\left(1+x_{ij}\right)\right)|\mathcal{H}_{0,T}\right]}{\exp\left(\widetilde{TS}_T\right)} \xrightarrow{P} 1
$$

uniformly in β.

Then, result (3.3) follows from

$$\frac{\ell_T(\theta,\beta)}{\exp\left(\widetilde{TS}_T\right)} =$$

$$\frac{\ell_T(\theta,\beta)}{E\left[\exp\left(TE_T\right)|\mathcal{H}_{0,T}\right]}\frac{E\left[\exp\left(TE_T\right)|\mathcal{H}_{0,T}\right]}{E\left[\exp\left(\widetilde{TS}_T+\sum_{i=1}^{B_N}\sum_{j=1}^{J}\ln\left(1+x_{ij}\right)\right)|\mathcal{H}_{0,T}\right]}$$

$$\times\frac{E\left[\exp\left(\widetilde{TS}_T+\sum_{i=1}^{B_N}\sum_{j=1}^{J}\ln\left(1+x_{ij}\right)\right)|\mathcal{H}_{0,T}\right]}{\exp\left(\widetilde{TS}_T\right)}.$$

Step 1. Using a Taylor expansion, we obtain

$$\begin{aligned}
&\left|\sum_{t=1}^{T}\left(l_t\left(\theta_T+\eta_t/T^{1/4}\right)-l_t\left(\theta_T\right)\right)\left(\eta_t\right)-\sum_{t=1}^{T}TE_t\right| \\
&\leq \sum_{t=1}^{T}\left\|l_t^{(5)}\left(\theta_T\right)\right\|\cdot M^5\cdot\frac{1}{T\sqrt[4]{T}} \\
&\leq \sup_{t,\theta\in\mathcal{N}}\left\|l_t^{(5)}\left(\theta\right)\right\|M^5\frac{1}{\sqrt[4]{T}} \\
&\leq const M^5\frac{1}{\sqrt[4]{T}} \\
&= o\left(1\right)
\end{aligned}$$

by Assumption 4. In the sequel, we will use sup instead of $\sup_{t,\theta\in\mathcal{N}}$ to simplify notation.

Step 2.

Let $TE_T=\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}TE_{it}$, $\widetilde{TS}_T=\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\widetilde{TS}_{it}$.

$$\begin{aligned}
TE_T-\widetilde{TS}_T &= \sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\left(TE_{it}-\widetilde{TS}_{it}\right) \\
&= \sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\left(TE_{it}-\widehat{TS}_{it}\right)+\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\left(\widehat{TS}_{it}-\widetilde{TS}_{it}\right)
\end{aligned}$$

where

$$\widehat{TS}_{it}=\frac{1}{2\sqrt{T}}E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)-\frac{1}{8T}\left[E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)\right]^2$$

In the sequel η_t is split in the following manner

$$\begin{aligned}
\eta_t &= \xi_t+\alpha_t, \\
\xi_t &= \eta_t-E(\eta_t|\mathcal{H}_{T_{i-1},T}), \\
\alpha_t &= E(\eta_t|\mathcal{H}_{T_{i-1},T}).
\end{aligned}$$

$\sum_t \widetilde{TS}_{it}$ can be decomposed as follows:

$$
\begin{aligned}
\sum_{t=T_{i-1}+1}^{T_i} \widehat{TS}_{it} &= \sum_{t=T_{i-1}+1}^{T_i} \widehat{TS}_{it}(\xi) + \sum_{t=T_{i-1}+1}^{T_i} \widehat{TS}_{it}(\alpha), \\
\sum_{t=T_{i-1}+1}^{T_i} \widehat{TS}_{it}(\xi) &= \frac{1}{2\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} E\left(m_{2,t}(\xi) \,|\mathcal{H}_{T_{i-1},T}\right) - \frac{1}{8T}\left[E\left(m_{2,t}(\xi)\,|\mathcal{H}_{T_{i-1},T}\right)\right]^2, \\
\sum_{t=T_{i-1}+1}^{T_i} \widehat{TS}_{it}(\alpha) &= \frac{1}{2\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} E\left(m_{2,t}(\alpha) \,|\mathcal{H}_{T_{i-1},T}\right) - \frac{1}{8T}\left[E\left(m_{2,t}(\alpha)\,|\mathcal{H}_{T_{i-1},T}\right)\right]^2.
\end{aligned}
$$

The mixed terms vanish because

$$E\left(\alpha_t \otimes \xi_t \mid \mathcal{H}_{T_{i-1},T}\right) = 0$$

Similarly, the Taylor Expansion in (7.14) can be rewritten as the sum of three parts, namely, the pure part w.r.t. ξ_t, the pure part w.r.t. α_t and the mixed part. That is,

$$TE_{it}(\xi_t) = \frac{1}{\sqrt[4]{T}} l_t^{(1)}(\xi_t) + \frac{1}{2\sqrt{T}} l_t^{(2)}(\xi_t,\xi_t) + \frac{1}{6\sqrt[4]{T^3}} l_t^{(3)}(\xi_t,\xi_t,\xi_t) + \frac{1}{24T} l_t^{(4)}(\xi_t,\xi_t,\xi_t,\xi_t),$$

$$TE_{it}(\alpha_t) = \frac{1}{\sqrt[4]{T}} l_t^{(1)}(\alpha_t) + \frac{1}{2\sqrt{T}} l_t^{(2)}(\alpha_t,\alpha_t) + \frac{1}{6\sqrt[4]{T^3}} l_t^{(3)}(\alpha_t,\alpha_t,\alpha_t) + \frac{1}{24T} l_t^{(4)}(\alpha_t,\alpha_t,\alpha_t,\alpha_t)$$

and

$$
\begin{aligned}
TE_{it}(\xi_t,\alpha_t) &= \frac{1}{2\sqrt{T}} \underbrace{l_t^{(2)}(\xi_t,\alpha_t)}_{\text{2 permutations}} + \frac{1}{6\sqrt[4]{T^3}} \underbrace{l_t^{(3)}(\xi_t,\xi_t,\alpha_t)}_{\text{3 permutations}} + \frac{1}{6\sqrt[4]{T^3}} \underbrace{l_t^{(3)}(\xi_t,\alpha_t,\alpha_t)}_{\text{3 permutations}} \\
&\quad + \frac{1}{24T} \underbrace{l_t^{(4)}(\xi_t,\xi_t,\xi_t,\alpha_t)}_{\text{4 permutations}} + \frac{1}{24T} \underbrace{l_t^{(4)}(\xi_t,\xi_t,\alpha_t,\alpha_t)}_{\text{6 permutations}} + \frac{1}{24T} \underbrace{l_t^{(4)}(\xi_t,\alpha_t,\alpha_t,\alpha_t)}_{\text{4 permutations}}
\end{aligned}
$$

Remark that for any linear function g and using the convention $\alpha_t = 0$ for $t > T$, we can write

$$
\begin{aligned}
& \sum_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} \left(g(\eta_t)\right) \\
= & \sum_{i=1}^{B_N} \left(\sum_{t=T_{i-1}+1}^{T_i} g(\xi_t) + \sum_{t=T_i+1}^{T_{i+1}} g(\alpha_t) \right)
\end{aligned}
$$

using $E(\eta_t|\mathcal{H}_{0,T}) = E(\eta_t) = 0$ with $\xi_t = \eta_t - E(\eta_t|\mathcal{H}_{T_{i-1},T})$ and $\alpha_t = E(\eta_t|\mathcal{H}_{T_i,T})$.

To summarize, we have the following decomposition

$$\begin{aligned}
&\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\left(TE_{it}-\widehat{TS}_{it}\right)\\
=\ &\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\left(TE_{it}\left(\xi_t\right)-\widehat{TS}_{it}\left(\xi_t\right)\right)\\
&+\sum_{i=1}^{B_N}\sum_{t=T_i+1}^{T_{i+1}}\left(TE_{it}\left(\alpha_t\right)-\widehat{TS}_{it}\left(\alpha_t\right)\right)\\
&+\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}TE_{it}\left(\xi_t,\alpha_t\right)
\end{aligned}$$

1) Term $\sum_{t=T_{i-1}+1}^{T_i}\left(\widehat{TS}_{it}-\widetilde{TS}_{it}\right)$:

$$\sum_{t=T_{i-1}+1}^{T_i}\left(\widehat{TS}_{it}-\widetilde{TS}_{it}\right)=\frac{1}{2\sqrt{T}}\sum_{t=T_{i-1}+1}^{T_i}\left[E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)-E\left(m_{2,t}|\mathcal{H}_{0,T}\right)\right] \quad \text{(T1)}$$

$$-\frac{1}{8T}\sum_{t=T_{i-1}+1}^{T_i}\left\{\left[E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)\right]^2-\left[E\left(m_{2,t}|\mathcal{H}_{0,T}\right)\right]^2\right\} \quad \text{(T2)}$$

The term (T2) converges to 0 uniformly in probability as shown below. Hence we have

$$\begin{aligned}
\sum_i\sum_{t=T_{i-1}+1}^{T_i}\left(\widehat{TS}_{it}-\widetilde{TS}_{it}\right) &= \frac{1}{2\sqrt{T}}\sum_i\sum_{t=T_{i-1}+1}^{T_i}\left[E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)-E\left(m_{2,t}|\mathcal{H}_{0,T}\right)\right]+o_p\left(1\right)\\
&= \frac{1}{2\sqrt{T}}\sum_i\sum_{t=T_i+1}^{T_{i+1}}\left[E\left(m_{2,t}|\mathcal{H}_{T_i,T}\right)-E\left(m_{2,t}|\mathcal{H}_{0,T}\right)\right]+o_p\left(1\right)
\end{aligned}$$

Hence $\sum_{i=1}^{B_N}\sum_{t=T_i+1}^{T_{i+1}}\left(\widehat{TS}_{it}\left(\xi\right)-\widetilde{TS}_{it}\left(\xi\right)-\ln(1+x\left(\xi\right))\right)\overset{P}{\to}0$ uniformly in β with $x\left(\xi\right)=\frac{1}{2\sqrt{T}}\left[E\left(m_{2,t}\left(\xi\right)|\mathcal{H}_{T_i,T}\right)-E\left(m_{2,t}\left(\xi\right)|\mathcal{H}_{0,T}\right)\right]$. The same is true for the term in α.

Now we show that (T2) converges to 0. We have

$$
\begin{aligned}
&\left\| \frac{1}{8T} \sum_{t=T_{i-1}+1}^{T_i} \left\{ \left[E\left(m_{2,t} | \mathcal{H}_{T_{i-1},T} \right) \right]^2 - \left[E\left(m_{2,t} | \mathcal{H}_{0,T} \right) \right]^2 \right\} \right\| \\
= &\left\| \frac{1}{8T} \sum_{t=T_{i-1}+1}^{T_i} \left\{ \left[E\left(m_{2,t} | \mathcal{H}_{T_{i-1},T} \right) \right] - \left[E\left(m_{2,t} | \mathcal{H}_{0,T} \right) \right] \right\} \left\{ \left[E\left(m_{2,t} | \mathcal{H}_{T_{i-1},T} \right) \right] + \left[E\left(m_{2,t} | \mathcal{H}_{0,T} \right) \right] \right\} \right\| \\
\leq & \frac{const}{8T} \sum_{t=T_{i-1}+1}^{T_i} \left\| \left[E\left(m_{2,t} | \mathcal{H}_{T_{i-1},T} \right) \right] - \left[E\left(m_{2,t} | \mathcal{H}_{0,T} \right) \right] \right\| \left(const + \| L_{t-1} \| \right) \\
\leq & \frac{const}{T} \sum_{t=T_{i-1}+1}^{T_i} \left(t - T_{i-1} \right) \lambda^{t-T_{i-1}} \\
\leq & \frac{const}{T}
\end{aligned}
$$

by the geometric ergodicity of η_t. Hence the term in $1/T$ is negligeable. As above, (T1) can be decomposed into a pure term in α_t and a pure term in ξ_t. Consider first the term in ξ_t. We use $|x - \log(1+x)| \leq x^2$. By Assumption 1, we have

$$
\begin{aligned}
& E \frac{1}{T} \left\{ \sum_{t=T_{i-1}+1}^{T_i} m_{2,t} \left[E(\xi_t \otimes \xi_t) - E(\xi_t \otimes \xi_t \mid \mathcal{H}_{T_{i-1},T}) \right] \right\}^2 \\
\leq \; & E \frac{1}{T} \left\{ \sum_{t=T_{i-1}+1}^{T_i} \| m_{2,t} \| \cdot \lambda^{t-T_{i-1}} \right\}^2 \\
= \; & E \frac{1}{T} \left\{ \sum_{t=T_{i-1}+1}^{T_i} \| m_{2,t} \| \cdot \sqrt{\lambda}^{t-T_{i-1}} \cdot \sqrt{\lambda}^{t-T_{i-1}} \right\}^2 \\
\leq \; & E \frac{1}{T} \left(\sum_{t=T_{i-1}+1}^{T_i} \| m_{2,t} \|^2 \cdot \lambda^{t-T_{i-1}} \right) \cdot \frac{1}{1-\lambda}
\end{aligned}
$$

Moreover we know,

$$
\| m_{2,t} \|^2 \leq const \cdot \left(\left\| l_t^{(1)} \right\|^4 + \left\| l_t^{(2)} \right\|^2 + \left\| l_t^{(1)} \right\|^2 \cdot \| L_{t-1} \|^2 \right)
$$

and

$$
\| L_{t-1} \| \leq (t - T_{i-1}) \cdot \left\| l_t^{(1)} \right\|
$$

So we have

$$
E \frac{1}{T} \left(\sum_{t=T_{i-1}+1}^{T_i} \| m_{2,t} \|^2 \cdot \lambda^{t-T_{i-1}} \right) \leq \frac{1}{T} \sum_{t=T_{i-1}+1}^{T_i} (const + const \cdot (t - T_{i-1})^2) \lambda^{t-T_{i-1}}
$$

Hence, for the sum over all blocks, we have

$$\begin{aligned}
& \sum_{i=1}^{B_N} \frac{1}{T} \left(\sum_{t=T_{i-1}+1}^{T_i} m_{2,t} \left[E(\xi_t \otimes \xi_t) - E(\xi_t \otimes \xi_t | \mathcal{H}_{T_{i-1},T}) \right] \right)^2 \\
\leq & \frac{T}{B_L} \frac{1}{T} \left(\frac{1}{1-\lambda} + \sum_{j=1}^{T_i - T_{i-1}} j^2 \lambda^j \right) \\
= & O(\frac{1}{B_L}) = o(1)
\end{aligned}$$

2) Term $TE_{it} - \widehat{TS}_{it}$:

We analyze the pure terms w.r.t. ξ_t. For simplicity, we drop ξ_t in the expressions. We use the notation $\sum\limits_{t=1}^{T} = \sum\limits_{i=1}^{B_N} \sum_{t=T_{i-1}+1}^{T_i} = \sum_i \sum_t$. Let $L_t = \sum_{s=T_{i-1}+1}^{t} l_t^{(1)}$ and

$$\begin{aligned}
L_1 &= \sum_t l_t^{(1)} \\
M_2 &= \sum_t m_{2,t} = \sum_t (l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)} + 2 l_t^{(1)} L_{t-1}) \\
M_3 &= \sum_t m_{3,t} = \sum_t (l_t^{(3)} + 3 l_t^{(2)} \otimes l_t^{(1)} + l_t^{(1)3}) \\
M_4 &= \sum_t m_{4,t} = \sum_t (l_t^{(4)} + 6 l_t^{(2)} \otimes l_t^{(1)2} + 4 l_t^{(3)} \otimes l_t^{(1)} + 3 l_t^{(2)} \otimes l_t^{(2)} + l_t^{(1)4})
\end{aligned}$$

where $l_t^{(1)}$, $m_{2,t}$, $m_{3,t}$, and $m_{4,t}$ come from the Bartlett identities described in Subsection

6. The pure terms $\sum_t TE_{it}(\xi_t)$ can be rewritten as follows

$$\frac{1}{\sqrt[4]{T}}L_1 - \frac{1}{2\sqrt{T}}L_1^2 + \frac{1}{3\sqrt[4]{T^3}}L_1^3 - \frac{1}{4T}L_1^4 \tag{P1}$$

$$+\frac{1}{2\sqrt{T}}M_2 - \frac{1}{8T}\sum_t m_{2,t}^2 \tag{P2}$$

$$-\frac{1}{2\sqrt[4]{T^3}}\sum_t l_t^{(1)} m_{2,t} \tag{P3}$$

$$+\frac{1}{2T}\sum_t l_t^{(1)2} m_{2,t} \tag{P4}$$

$$+\frac{1}{6\sqrt[4]{T^3}}M_3 \tag{P5}$$

$$-\frac{1}{6T}\sum_t m_{3,t} l_t^{(1)} \tag{P6}$$

$$-\frac{1}{\sqrt[4]{T^3}}\sum_t l_t^{(1)} L_{t-1}^2 \tag{P7}$$

$$+\frac{1}{T}\sum_t l_t^{(1)2} L_{t-1}^2 \tag{P8}$$

$$+\frac{1}{T}\sum_t l_t^{(1)} L_{t-1}^3 \tag{P9}$$

$$+\frac{1}{2T}\sum_t l_t^{(1)} L_{t-1} m_{2,t} \tag{P10}$$

$$+\frac{1}{24T}M_4 \tag{P11}$$

And we add to (P2) the term $-\sum_t \widehat{TS}_{it}(\xi_t)$:

$$-\frac{1}{2\sqrt{T}}\sum_t E(m_{2,t}|\mathcal{H}_{T_{i-1},T}) + \frac{1}{8T}\sum_t \left[E(m_{2,t}|\mathcal{H}_{T_{i-1},T})\right]^2 \tag{7.15}$$

Now we show that each term (P1) to (P10) can be approximated by a term of the form $\ln(1+x)$. We also show that the sum over the blocks of (P11) goes to zero and therefore can be neglected.

(P1):
Using a Taylor expansion we have

$$\left|(P1) - \ln(1 + \frac{1}{\sqrt[4]{T}}L_1)\right| \leq const \cdot \frac{1}{T\sqrt[4]{T}}\left\|L_1\right\|^5$$

Now we analyze the moment conditions needed.

$$E\left(\left\|\sum_t l_t^{(1)}\right\|^5\right) = \left[\sqrt[5]{E\left(\left\|\sum_t l_t^{(1)}\right\|^5\right)}\right]^5 \leq \left[\sum_t \sqrt[5]{E\left\|l_t^{(1)}\right\|^5}\right]^5 = B_L^5 \cdot \left[\frac{1}{B_L}\sum_t \sqrt[5]{E\left\|l_t^{(1)}\right\|^5}\right]^5$$

by the triangular inequality. This term is $O\left(B_L^5\right)$ provided $\sup E\left\|l_t^{(1)}\right\|^5 < \infty$. Using the fact that if $X_T > 0$, $EX_T \to 0$ implies that $X_T \xrightarrow{P} 0$, the sum over all blocks goes to zero if the following condition holds:

$$\frac{T}{B_L} \cdot \frac{1}{T\sqrt[4]{T}} \cdot B_L^5 = o(1)$$

Then we get the bound condition

$$B_L = o(T^{\frac{1}{16}}), \tag{B1}$$

which is assumed to hold throughout this proof.

Consider term (P2)+(7.15):

$$(P2)+(7.15) = \frac{1}{2\sqrt{T}}M_2 - \frac{1}{8T}\sum_t m_{2,t}^2 - \frac{1}{2\sqrt{T}}\sum_t E(m_{2,t}|\mathcal{H}_{T_{i-1},T}) + \frac{1}{8T}\sum_t \left[E(m_{2,t}|\mathcal{H}_{T_{i-1},T})\right]^2$$

We want to show that (P2)+(7.15) can be approximated by

$$\ln\left(1+x_{2t}\right) \equiv \log(1 + \frac{M_2 - E(M_2|\mathcal{H}_{T_{i-1},T})}{2\sqrt{T}} + \frac{K}{T})$$

where

$$K = \frac{1}{8}\left\{\left[M_2 - E(M_2|\mathcal{H}_{T_{i-1},T})\right]^2 - \sum_t \left[m_{2,t}^2 - \left[E(m_{2,t}|\mathcal{H}_{T_{i-1},T})\right]^2\right]\right\}$$

For arbitrary A and B, a Taylor expansion gives:

$$\begin{aligned}
&\left|\log\left(1 + \frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{8T}\right) - \left(\frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{8T}\right) + \frac{1}{2}\left(\frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{8T}\right)^2\right| \\
\leq\ & \frac{1}{3}\left|\frac{A}{\sqrt{T}} + \frac{A^2}{2T} - \frac{B}{8T}\right|^3
\end{aligned}$$

Denote $C = \frac{A^2}{2} - \frac{B}{8}$, then we have

$$\left|\log(1 + \frac{A}{\sqrt{T}} + \frac{C}{T}) - (\frac{A}{\sqrt{T}} - \frac{B}{8T})\right|$$

$$\leq \left|-\frac{A^2}{2T} + \frac{1}{2}(\frac{A}{\sqrt{T}} + \frac{C}{T})^2\right| + \frac{1}{3}\left|\frac{A}{\sqrt{T}} + \frac{C}{T}\right|^3 \tag{7.16}$$

$$= \left|\frac{1}{2}\frac{C}{T^2}^2 + \frac{AC}{T\sqrt{T}}\right| + \frac{1}{3}\left|\frac{A}{\sqrt{T}} + \frac{C}{T}\right|^3 \tag{7.17}$$

$$\leq const \cdot \left(\left\|\frac{A}{\sqrt{T}}\right\| \cdot \frac{\|C\|}{T} + \frac{\|C\|^2}{T^2} + \frac{\|A\|^3}{T\sqrt{T}} + \frac{\|C\|^3}{T^3}\right) \tag{7.18}$$

We apply this result to $A = M_2 - E(M_2|\mathcal{H}_{T_{i-1},T})$ and $C = K$. We want to establish that the expectation of the r.h.s. of (7.18) goes to zero uniformly.

First we analyze $\left\|\frac{A}{\sqrt{T}}\right\|^3$.

$$\begin{aligned}\left\|\frac{A}{\sqrt{T}}\right\|^3 &\leq const \cdot \left[\frac{1}{T^{3/2}}\left\|M_2\right\|^3 + \frac{1}{T^{3/2}}E^3(\|M_2\| \mid \mathcal{H}_{T_{i-1},T})\right] \\ &\leq const \cdot \left[\frac{1}{T^{3/2}}\left\|M_2\right\|^3 + \frac{1}{T^{3/2}}E(\|M_2\|^3 \mid \mathcal{H}_{T_{i-1},T})\right]\end{aligned}$$

where the first inequality follows from Lemma 7.4 and the second inequality comes from Jensen's Inequality as the function $f(x) = x^3$ is convex in $\mathcal{R}^+$.

Then

$$\begin{aligned}E\left\|\frac{A}{\sqrt{T}}\right\|^3 &\leq const \cdot \frac{1}{T^{3/2}}E\left\|M_2\right\|^3 \leq const \cdot \frac{1}{T^{3/2}}E\left(\sum_t \|m_{2,t}\|\right)^3 \\ &\leq const \cdot \frac{1}{T^{3/2}}B_L^2 \cdot E\left(\sum_t \|m_{2,t}\|^3\right) \\ &\leq const \cdot \frac{1}{T^{3/2}}B_L^2 \cdot E\sum_t\left(\left\|l_t^{(2)}\right\|^3 + \left\|l_t^{(1)}\right\|^6 + \left\|l_t^{(1)}\right\|^3 \cdot \|L_{t-1}\|^3\right) \\ &= O\left(\frac{B_L^6}{T^{3/2}}\right)\end{aligned}$$

provided $\sup E\left\|l_t^{(2)}\right\|^3 < \infty$ and $\sup E\left\|l_t^{(1)}\right\|^6 < \infty$. Hence

$$\sum_i E\left\|\frac{A}{\sqrt{T}}\right\|^3 = O\left(\frac{T}{B_L}\frac{B_L^6}{T^{3/2}}\right) = O\left(\frac{B_L^5}{T^{1/2}}\right) = o(1).$$

Moreover, from Holder's inequality, we have

$$E\left\|\frac{A}{\sqrt{T}}\right\|^2 \leq \left[E\left(\left\|\frac{A}{\sqrt{T}}\right\|^3\right)\right]^{2/3} = O\left(\frac{B_L^4}{T}\right)$$

Now we analyze the term $\|C\|^3$.

$$C = \frac{A^2}{2} - \frac{B}{8} = \frac{A^2}{2} - \frac{1}{8}\sum_t\left[m_{2,t}^2 - \left[E(m_{2,t}|\mathcal{H}_{T_{i-1},T})\right]^2\right]$$

and

$$\begin{aligned}\|A\|^2 &\leq const \cdot \left[\|M_2\|^2 + E^2(\|M_2\| \mid \mathcal{H}_{T_{i-1},T})\right] \\ &\leq const \cdot \left[\|M_2\|^2 + E(\|M_2\|^2 \mid \mathcal{H}_{T_{i-1},T})\right]\end{aligned}$$

Again, the second inequality comes from Jensen's Inequality.

Then we have

$$\begin{aligned}
\|C\|^3 &\le const \cdot \left(\|M_2\|^2 + E(\|M_2\|^2 \,|\mathcal{H}_{T_{i-1},T}) + \sum_t \left[\|m_{2,t}\|^2 + \left[E(\|m_{2,t}\| \,|\mathcal{H}_{T_{i-1},T}) \right]^2 \right] \right)^3 \\
&\le const \cdot \left((\sum_t \|m_{2,t}\|)^2 + E((\sum_t \|m_{2,t}\|)^2 \mid \mathcal{H}_{T_{i-1},T}) + \sum_t \left[\|m_{2,t}\|^2 + E(\|m_{2,t}\|^2 \,|\mathcal{H}_{T_{i-1},T}) \right] \right)^3 \\
&\le const \cdot (B_L+1)^3 \left(\sum_t \left[\|m_{2,t}\|^2 + E(\|m_{2,t}\|^2 \,|\mathcal{H}_{T_{i-1},T}) \right] \right)^3 \\
&\le const \cdot (B_L+1)^3 \cdot B_L^2 \sum_t \left[\|m_{2,t}\|^6 + E(\|m_{2,t}\|^6 \,|\mathcal{H}_{T_{i-1},T}) \right]
\end{aligned}$$

Therefore,

$$\begin{aligned}
E\|C\|^3 &\le const\ (B_L+1)^3 B_L^2 \sum_t E\|m_{2,t}\|^6 \\
&\le const\ \frac{(B_L+1)^3}{T^3} B_L^2 \sum_t E\left(\left\| l_t^{(2)} \right\|^6 + \left\| l_t^{(1)} \right\|^{12} + \left\| l_t^{(1)} \right\|^6 \cdot \|L_{t-1}\|^6 \right) \\
&= O\left(B_L^{12}\right)
\end{aligned}$$

provided $\sup E \left\| l_t^{(2)} \right\|^6 < \infty$ and $\sup E \left\| l_t^{(1)} \right\|^{12} < \infty$. Hence

$$\sum_i \frac{E\|C\|^3}{T^3} = O\left(\frac{T}{B_L} \frac{1}{T^3} B_L^{12} \right) = o(1).$$

Moreover by Holder's inequality

$$E\|C\|^2 \le \left(E\|C\|^3 \right)^{2/3} = O\left(B_L^8\right).$$

Hence

$$\sum_i \frac{E\|C\|^2}{T^2} = O\left(\frac{T}{B_L} \frac{1}{T^2} B_L^8 \right) = O\left(\frac{B_L^7}{T} \right) = o(1).$$

Now we analyze term $\left\| \frac{A}{\sqrt{T}} \right\| \cdot \|C\|$.

$$\begin{aligned}
E\left(\left\| \frac{A}{\sqrt{T}} \right\| \cdot \frac{\|C\|}{T} \right) &\le \sqrt{ E\left\| \frac{A}{\sqrt{T}} \right\|^2 \cdot E\frac{\|C\|^2}{T^2} } \\
&= O\left(\frac{B_L^4}{T^{3/2}} \right).
\end{aligned}$$

Hence

$$\sum_i E\left(\left\|\frac{A}{\sqrt{T}}\right\| \cdot \frac{\|C\|}{T}\right) = O\left(\frac{B_L^3}{T^{1/2}}\right) = o(1).$$

Now we consider the terms (P3) to (P10). Remark that (P3) to (P10) correspond to $\widetilde{x}_{i3}$ to $\widetilde{x}_{i10}$ in Lemma 7.7. From the Taylor expansion, we have

$$|x - \log(1+x)| \leq const \cdot x^2$$

We need to show that $E\left\|x^2\right\|$ converges to zero uniformly in β. To do so, we use the bounds given by Lemma 7.7. By Holder's inequality, $E\left(\|x\|^2\right) \leq E\left(\|x\|^4\right)^{2/4}$.

Term (P3):

$$\begin{aligned}
E\left(\|P_3\|^2\right) &= \frac{1}{T^{3/2}} E\left(\|x_{i3}\|^2\right) \\
&\leq \frac{1}{T^{3/2}}\left(B_L^8\right)^{1/2} = \frac{B_L^4}{T^{3/2}}, \\
\sum_i E\left(\|P_3\|^2\right) &\leq \frac{T}{B_L}\frac{B_L^4}{T^{3/2}} = \frac{B_L^3}{T^{1/2}}.
\end{aligned}$$

Term (P4):

$$\begin{aligned}
E\left(\|P_4\|^2\right) &\leq \frac{1}{T^2}\left(B_L^8\right)^{1/2} = \frac{B_L^4}{T^2}, \\
\sum_i E\left(\|P_4\|^2\right) &\leq \frac{T}{B_L}\frac{B_L^4}{T^2} = \frac{B_L^3}{T}.
\end{aligned}$$

Term (P5):

$$\begin{aligned}
E\left(\|P_5\|^2\right) &\leq \frac{1}{T^{3/2}}\left(B_L^4\right)^{1/2} = \frac{B_L^2}{T^{3/2}}, \\
\sum_i E\left(\|P_5\|^2\right) &\leq \frac{B_L}{T^{1/2}}.
\end{aligned}$$

Term (P6):

$$\begin{aligned}
E\left(\|P_6\|^2\right) &\leq \frac{1}{T^2}\left(B_L^8\right)^{1/2} = \frac{B_L^2}{T^{3/2}}, \\
\sum_i E\left(\|P_6\|^2\right) &\leq \frac{B_L}{T^{1/2}}.
\end{aligned}$$

Term (P7):

$$
\begin{aligned}
E\left(\|P_7\|^2\right) &\le \frac{1}{T^{3/2}}\left(B_L^{12}\right)^{1/2} = \frac{B_L^6}{T^{3/2}}, \\
\sum_i E\left(\|P_7\|^2\right) &\le \frac{B_L^5}{T^{1/2}}.
\end{aligned}
$$

Term (P8):

$$
\begin{aligned}
E\left(\|P_8\|^2\right) &\le \frac{1}{T^2}\left(B_L^{12}\right)^{1/2} = \frac{B_L^6}{T^2}, \\
\sum_i E\left(\|P_8\|^2\right) &\le \frac{B_L^5}{T}.
\end{aligned}
$$

Term (P9):

$$
\begin{aligned}
E\left(\|P_9\|^2\right) &\le \frac{1}{T^2}\left(B_L^{16}\right)^{1/2} = \frac{B_L^8}{T^2}, \\
\sum_i E\left(\|P_9\|^2\right) &\le \frac{B_L^7}{T}.
\end{aligned}
$$

Term (P10):

$$
\begin{aligned}
E\left(\|P_{10}\|^2\right) &\le \frac{1}{T^2}\left(B_L^{16}\right)^{1/2} = \frac{B_L^8}{T^2}, \\
\sum_i E\left(\|P_{10}\|^2\right) &\le \frac{B_L^7}{T}.
\end{aligned}
$$

Term (P11):

$$
\left\|\frac{1}{T}\sum_{t=1}^{T} m_{4,t}\left(\xi_t,\xi_t,\xi_t,\xi_t\right)\right\| \le \left\|\frac{1}{T}\sum_{t=1}^{T} m_{4,t}\right\| M^4
$$

because $\|\xi_t\| \le M$.

$$
\begin{aligned}
E\|m_{4,t}\| &\le E\left\|l_t^{(4)} + 6l_t^{(2)} \otimes l_t^{(1)2} + 4l_t^{(3)} \otimes l_t^{(1)} + 3l_t^{(2)} \otimes l_t^{(2)} + l_t^{(1)4}\right\| \\
&\le E\left(\left\|l_t^{(4)}\right\| + 6\left\|l_t^{(2)}\right\|\left\|l_t^{(1)2}\right\| + 4\left\|l_t^{(3)}\right\|\left\|l_t^{(1)}\right\| + 3\left\|l_t^{(2)}\right\|^2 + \left\|l_t^{(1)4}\right\|\right) \\
&< \infty
\end{aligned}
$$

provided $\sup E\left\|l_t^{(4)}\right\| < \infty$, $\sup E\left\|l_t^{(3)}\right\|^2 < \infty$, $\sup E\left\|l_t^{(2)}\right\|^2 < \infty$, $\sup E\left\|l_t^{(1)}\right\|^4 < \infty$. As $m_{4,t}$ is a martingale, $\frac{1}{T}\sum m_{4,t} = o_p(1)$ and hence

$$
\frac{1}{T}\|M_4\| = o_p(1)
$$

uniformly in β.

Hence we have shown that $\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\left(TE_{it}(\xi_t)-\widehat{TS}_{it}(\xi_t)-\sum_{j=1}^{4}\ln(1+x_{ijt})\right)\xrightarrow{P}0$ uniformly in β.

We analyze the pure terms w.r.t. α_t. We can do the same approximation as for the terms in ξ_t using the property that for $T_{i+1}\geq t>T_i$

$$\begin{aligned}\|\alpha_t\| &= \|E(\eta_t|\mathcal{H}_{T_i,T})\|\\ &\leq \lambda^{t-T_i}\end{aligned}$$

by the property of geometric ergodicity of η_t. Note that all the terms in $1/T$ can be neglected. For illustration, we treat the case of term (P9). Remark that

$$\begin{aligned}\|L_{t-1}\|^3 &= \left\|\sum_{t=T_i+1}^{t-1} l_s^{(1)}(\alpha_s)\right\|^3\\ &\leq M^3\sup\left\|l_t^{(1)}\right\|^3(t-T_i)^3\end{aligned}$$

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}L_{t-1}^3\right\| &\leq M^3\sup\left\|l_t^{(1)}\right\|^4\frac{1}{T}\sum_{t=T_i+1}^{T_{i+1}}(t-T_i)^3\|\alpha_t\|\\ &\leq const\frac{1}{T}\sum_{t=T_i+1}^{T_{i+1}}(t-T_i)^3\lambda^{t-T_i}\\ &\leq const\frac{1}{T}.\end{aligned}$$

Hence

$$\left\|\frac{1}{T}\sum_i\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}L_{t-1}^3\right\|\leq\frac{T}{B_L}\frac{1}{T}\to 0.$$

Therefore the term (P9) is negligeable. And so are the other terms in $1/T$. The remaining terms in α are (P1), (P2), (P3), (P5) and (P7).

Mixed terms:

The mixed terms are as follows:

$$\frac{1}{2\sqrt{T}}\sum\underbrace{l_t^{(2)}(\xi_t,\alpha_t)}_{\text{2 terms}}+\frac{1}{6\sqrt[4]{T^3}}\sum\underbrace{l_t^{(3)}(\xi_t,\xi_t,\alpha_t)}_{\text{3 terms}}+\frac{1}{6\sqrt[4]{T^3}}\sum\underbrace{l_t^{(3)}(\xi_t,\alpha_t,\alpha_t)}_{\text{3 terms}}$$
$$+\frac{1}{24T}\sum\underbrace{l_t^{(4)}(\xi_t,\xi_t,\xi_t,\alpha_t)}_{\text{4 terms}}+\frac{1}{24T}\sum\underbrace{l_t^{(4)}(\xi_t,\xi_t,\alpha_t,\alpha_t)}_{\text{6 terms}}+\frac{1}{24T}\sum\underbrace{l_t^{(4)}(\xi_t,\alpha_t,\alpha_t,\alpha_t)}_{\text{4 terms}}$$

For all the 4th-order terms,

$$\sum_i E\frac{1}{T}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(4)}(\cdot,\cdot,\cdot,\alpha_t) \le \frac{T}{BL}\frac{1}{T}\sup E\left\|l_t^{(4)}\right\| \cdot M^3 \cdot \frac{1}{1-\lambda}$$

which converges to zero uniformly with the moment condition

$$\sup E\left\|l_t^{(4)}\right\| < \infty$$

For the 3rd-order terms, we apply the Bartlett Identity,

$$M^{(3)}(a,b,c) = l^{(3)}(a,b,c)+l^{(2)}(a,b)l^{(1)}(c)+l^{(2)}(a,c)l^{(1)}(b)+l^{(2)}(b,c)l^{(1)}(a)+l^{(1)}(a)l^{(1)}(b)l^{(1)}(c)$$

Hence the mixed-terms can be written as

$$\frac{1}{2\sqrt{T}}\sum\left(l_t^{(2)}(\alpha_t,\xi_t)+l_t^{(2)}(\xi_t,\alpha_t)\right) \tag{R1}$$

$$+\frac{1}{6\sqrt[4]{T^3}}\sum\left(\underbrace{M_t^{(3)}(\alpha_t,\alpha_t,\xi_t)}_{\text{3 terms}}+\underbrace{M_t^{(3)}(\alpha_t,\xi_t,\xi_t)}_{\text{3 terms}}\right) \tag{R2}$$

$$-\frac{1}{6\sqrt[4]{T^3}}\sum\left(3l_t^{(2)}(\alpha_t,\xi_t)l_t^{(1)}(\alpha_t)+3l_t^{(2)}(\xi_t,\alpha_t)l_t^{(1)}(\alpha_t)+3l_t^{(2)}(\alpha_t,\alpha_t)l_t^{(1)}(\xi_t)\right) \tag{R3}$$

$$-\frac{1}{6\sqrt[4]{T^3}}\sum\left(3l_t^{(2)}(\alpha_t,\xi_t)l_t^{(1)}(\xi_t)+3l_t^{(2)}(\xi_t,\alpha_t)l_t^{(1)}(\xi_t)\right) \tag{R4}$$

$$-\frac{1}{6\sqrt[4]{T^3}}\sum 3l_t^{(1)2}(\alpha_t)l_t^{(1)}(\xi_t) \tag{R5}$$

$$-\frac{1}{6\sqrt[4]{T^3}}\sum\left(3l_t^{(1)2}(\xi_t)l_t^{(1)}(\alpha_t)+3l_t^{(2)}(\xi_t,\xi_t)l_t^{(1)}(\alpha_t)\right) \tag{7.19}$$

The term (7.19) is rewritten in the following way.

$$-\frac{1}{2\sqrt[4]{T^3}}\sum l_t^{(1)}(\alpha_t)m_{2,t}(\xi_t,\xi_t) \tag{7.20}$$

$$+\frac{1}{\sqrt[4]{T^3}}\sum l_t^{(1)}(\alpha_t)l_t^{(1)}(\xi_t)L_{t-1}(\xi_t) \tag{R6}$$

Moreover, we have

$$
\begin{aligned}
& -\frac{1}{2\sqrt[4]{T^3}}\sum_i \sum_{t=T_{i-1}+1}^{T_i} l_t^{(1)}(\alpha_t) m_{2,t}(\xi_t,\xi_t) \\
= & -\frac{1}{2\sqrt[4]{T^3}}\sum_i \sum_{t=T_{i-1}+1}^{T_i} l_t^{(1)}(\alpha_t) m_{2,t}\left[(\xi_t,\xi_t) - E\left[(\xi_t\otimes\xi_t)|\mathcal{H}_{T_{i-1},T}\right]\right] \qquad \text{(R7)} \\
& -\frac{1}{2\sqrt[4]{T^3}}\sum_i \sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}(\alpha_t) m_{2,t}\left[E\left[(\xi_t\otimes\xi_t)|\mathcal{H}_{T_i,T}\right] - E\left[(\xi_t\otimes\xi_t)\right]\right] \qquad \text{(R8)} \\
& -\frac{1}{2\sqrt[4]{T^3}}\sum_i \sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}(\alpha_t) m_{2,t} E\left[(\xi_t\otimes\xi_t)\right] \qquad \text{(R9)}
\end{aligned}
$$

Note that the sum in (R8) and (R9) is over T_i+1 to T_{i+1}, this follows from a simple change of indice (replace i by $i+1$). Each term (R1) to (R9) (denoted x for convenience) can be approximated by terms $\ln(1+x)$. The terms $E(x^2)$ involve α_t and hence their sums converge to 0 uniformly in β.

Step 3.

As $\exp(TS_T)$ is $\mathcal{H}_{0,T}$−measurable, we have

$$
\frac{E\left[\exp\left(TS_T + \sum_{i=1}^{T_B}\sum_j \ln(1+x_{ij})\right)|\mathcal{H}_{0,T}\right]}{\exp(TS_T)} = E\left[\prod_{i=1}^{T_B}\prod_{j=1}^{J}(1+x_{ij})\,|\mathcal{H}_{0,T}\right]
$$

J is equal to 23, because there are 11 pure terms in ξ_t, 6 pure terms in α_t, and 6 mixed terms. The product can be rewritten as $\prod_{j=1}^{J}(1+x_{ij}) = 1 + \sum_{j=1}^{J} x_{ij} + \sum_{j\neq l} x_{ij}x_{il} + \sum_{j\neq l\neq p} x_{ij}x_{il}x_{ip} + ... + \prod_{j=1}^{J} x_{ij}$ where each x_{ij} is of the form $\frac{1}{T^{\alpha_j}}\sum_t x_{ijt}$. Hence $\prod_{j=1}^{J}(1+x_{ij})$ is 1 plus a sum of terms of the form $\prod_{j\in d} x_{ij}$, where d is a partition of $1,2,...,J$. Each of these terms can be treated individually. We need to compute $\Delta_{i,T}$ and check Conditions (7.4) and (7.5) in Lemma 7.2.

Consider the case $\sum_{j\in d}\alpha_j > 1$. Note that as soon as there are four terms, we have $\sum_{j\in d}\alpha_j \geq 1.5$ (d is a partition of $1,2,...,J$ with cardinal 4). By Lemma 7.7, we have

$$
E\left(\left\|\sum_t \tilde{x}_{ijt}\right\|^4\right) \leq const B_L^{16}.
$$

Hence by Lemma 7.6, we have

$$
\sum_i E\left(\prod_{j\in d}\|x_{ij}\|\right) \leq \frac{T}{B_L}\frac{B_L^{16}}{T^{3/2}} = \frac{B_L^{15}}{T^{1/2}} = o(1)
$$

for $B_L = o\left(T^{1/30}\right)$. For this choice of B_L, the conditions (7.4) and (7.5) are satisfied. If there are more than four terms, the conditions (7.4) and (7.5) are again satisfied. Indeed by Lemma 7.7 and Holder inequality, we have

$$E\left(\|x_{ij}\|\right) \leq const\frac{B_L^4}{T^{\alpha_j}} = o\left(1\right).$$

As $\|\alpha_t\|$ and $\|\xi_t\|$ are bounded by M, there is an $\mathcal{H}_{0,T}-$ measurable set A_T, such that $\|x_{ij}\| \leq 1/2$ on A_T and $P\left(A_T\right) \to 1$. Hence

$$\begin{aligned} \|\Delta_i\| &= E\left[\left\|\prod_{j\in d_1} x_{ij} \prod_{k\in d_1} x_{ik}\right\| |\mathcal{H}_{T_{i-1},T}\right] \\ &\leq \frac{1}{2}E\left[\left\|\prod_{j\in d_1} x_{ij}\right\| |\mathcal{H}_{T_{i-1},T}\right]. \end{aligned}$$

And the result follows from above.

In the case where there are fewer than 4 terms but $\sum_{j\in d}\alpha_j > 1$, Lemma 7.6 shows that the conditions (7.4) and (7.5) are also satisfied. This takes care of all the terms for which $\sum_{j\in d}\alpha_j > 1$. The terms with $\sum_{j\in d}\alpha_j \leq 1$ are treated on a case by case basis below.

1) Pure terms in ξ_t

The x_{ij} correspond to $P_1, P_2, ..., P_{10}$, and T_1 :

$$\begin{aligned} x_{i1} &= \frac{L_1}{T^{1/4}}, \\ x_{i2} &= x_{i20} + x_{i21} \\ x_{i20} &= \frac{M_2 - E\left(M_2|\mathcal{H}_{T_{i-1},T}\right)}{2\sqrt{T}}, \\ x_{i21} &= \frac{1}{8T}\left[M_2 - E\left(M_2|\mathcal{H}_{T_{i-1},T}\right)\right]^2 - \frac{1}{8T}\left[\sum_t m_{2,t}^2 - \sum_t \left[E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)\right]^2\right]. \end{aligned}$$

$$
\begin{aligned}
x_{i3} &= -\frac{1}{2\sqrt[4]{T^3}}\sum_t l_t^{(1)} m_{2,t}, \\
x_{i4} &= \frac{1}{2T}\sum_t l_t^{(1)2} m_{2,t}, \\
x_{i5} &= \frac{1}{6\sqrt[4]{T^3}} M_3, \\
x_{i6} &= -\frac{1}{6T}\sum_t l_t^{(1)} m_{3,t}, \\
x_{i7} &= -\frac{1}{\sqrt[4]{T^3}}\sum_t l_t^{(1)} L_{t-1}^2, \\
x_{i8} &= \frac{1}{T}\sum_t l_t^{(1)2} L_{t-1}^2, \\
x_{i9} &= \frac{1}{T}\sum_t l_t^{(1)} L_{t-1}^3, \\
x_{i10} &= \frac{1}{2T}\sum_t l_t^{(1)} L_{t-1} m_{2,t}, \\
x_{i11} &= \frac{1}{2\sqrt{T}}\sum_{t=T_i+1}^{T_{i+1}} \left[E\left(m_{2,t}|\mathcal{H}_{T_i,T}\right) - E\left(m_{2,t}|\mathcal{H}_{0,T}\right)\right]
\end{aligned}
$$

Note that x_{i11} is the only term for which the sum is over $T_i + 1$ to T_{i+1}.

(a) Terms for which $\sum_j \alpha_j = 1$.
Here is the list of such terms:

$$
\begin{aligned}
& x_{i21}, \\
& x_{i4} + x_{i10} + x_{i1}x_{i3}, \\
& x_{i6} + x_{i1}x_{i5}, \\
& x_{i8} + x_{i9} + x_{i1}x_{i7}.
\end{aligned}
$$

Terms x_{i21} :

$$
\begin{aligned}
& \Delta_i \\
= \; & E\left(x_{i21}|\mathcal{H}_{T_{i-1},T}\right) \\
= \; & \frac{1}{8T}\left\{E\left(M_2^2|\mathcal{H}_{T_{i-1},T}\right) - E\left(M_2|\mathcal{H}_{T_{i-1},T}\right)^2 - \left[\sum_t E\left(m_{2,t}^2|\mathcal{H}_{T_{i-1},T}\right) - \sum_t \left[E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right)\right]^2\right]\right\} \\
= \; & \frac{1}{8T}\left\{\sum_{t\neq s} E\left(m_{2,t}m_{2,s}|\mathcal{H}_{T_{i-1},T}\right) - \sum_{t\neq s} E\left(m_{2,t}|\mathcal{H}_{T_{i-1},T}\right) E\left(m_{2,s}|\mathcal{H}_{T_{i-1},T}\right)\right\}.
\end{aligned}
$$

Δ_i is a martingale in t for $t > s$ and in s for $s > t$. It is easy to show that $\sum_i E\left(\Delta_i^2\right) \to 0$.

Term $x_{i4} + x_{i10} + x_{i1}x_{i3}$:

$$
\begin{aligned}
& x_{i4} + x_{i10} + x_{i1}x_{i3} \\
= \; & \frac{1}{2T}\sum_t l_t^{(1)2} m_{2,t} + \frac{1}{2T}\sum_t l_t^{(1)} L_{t-1} m_{2,t} - \frac{1}{2T}\sum_{t,s} l_s^{(1)} l_t^{(1)} m_{2,t} \\
= \; & -\frac{1}{2T}\sum_{t,s>t} l_s^{(1)} l_t^{(1)} m_{2,t}
\end{aligned}
$$

$$
\Delta_i = -\frac{1}{2T}\sum_{s,t<s} l_s^{(1)} l_t^{(1)} m_{2,t} E\left(\xi_s \xi_t^2 | \mathcal{H}_{T_{i-1},T}\right)
$$

is a martingale in s. Using the fact that $\xi_t^2 \le 4M^2$, we have

$$
\begin{aligned}
E\left(\Delta_i^2\right) &= \frac{1}{4T^2}\sum_s E\left[l_s^{(1)2}\left(\sum_{t<s} l_t^{(1)} m_{2,t}\right)^2\right] E\left[E\left(\xi_s\xi_t^2|\mathcal{H}_{T_{i-1},T}\right)^2\right] \\
&\le \frac{M^2}{T^2}\sum_s \left[E\left(l_s^{(1)4}\right) E\left(\left(\sum_{t<s} l_t^{(1)} m_{2,t}\right)^4\right)\right]^{1/2} E\left[E\left(\xi_s|\mathcal{H}_{T_{i-1},T}\right)^2\right] \\
&\le const\frac{B_L^2}{T^2}\sum_s \lambda^{2(s-T_{i-1})}.
\end{aligned}
$$

Hence

$$
\sum_i E\left(\Delta_i^2\right) \sim \frac{B_L}{T} \to 0.
$$

Term $x_{i6} + x_{i1}x_{i5}$:

$$
\begin{aligned}
x_{i6} + x_{i1}x_{i5} &= -\frac{1}{6T}\sum_t l_t^{(1)} m_{3,t} + \frac{1}{6T}\sum_{t,s} l_t^{(1)} m_{3,s} \\
&= \frac{1}{6T}\sum_{t>s} l_t^{(1)} m_{3,s} + \frac{1}{6T}\sum_{t<s} l_t^{(1)} m_{3,s}.
\end{aligned}
$$

We can treat separately the two terms on the r.h.s. They are both martingales. We get the same rate as for the previous case.

Term $x_{i8} + x_{i9} + x_{i1}x_{i7}$:

$$
\begin{aligned}
& x_{i8} + x_{i9} + x_{i1}x_{i7} \\
= \; & \frac{1}{T}\sum_t l_t^{(1)2} L_{t-1}^2 + \frac{1}{T}\sum_t l_t^{(1)} L_{t-1}^3 - \frac{1}{T}\sum_t l_t^{(1)} L_{t-1}^2 \sum_s l_s^{(1)} \\
= \; & -\frac{1}{T}\sum_s l_s^{(1)} \sum_{t<s} l_t^{(1)} L_{t-1}^2
\end{aligned}
$$

Δ_i is again a martingale, we obtain

$$\begin{aligned} E\left(\Delta_i^2\right) &\leq const\frac{1}{T^2}\sum_s E\left[l_s^{(1)2}\left(\sum_{t<s} l_t^{(1)} L_{t-1}^2\right)^2\right]\lambda^{2(s-T_{i-1})} \\ &\leq const\frac{1}{T^2}\sum_s\left[E\left(l_s^{(1)4}\right)E\left(\left(\sum_{t<s} l_t^{(1)} L_{t-1}^2\right)^4\right)\right]^{1/2}\lambda^{2(s-T_{i-1})} \\ &\leq const\frac{B_L^6}{T^2} \end{aligned}$$

Hence

$$\sum_i E\left(\Delta_i^2\right) \sim \frac{B_L^5}{T} \to 0.$$

(b) Terms for which $\sum_j \alpha_j < 1$.
The list of such terms is

$$\begin{aligned} &x_{i1}, \\ &x_{i20}, \\ &x_{i3} + x_{i1}x_{i20}, \\ &x_{i5}, \\ &x_{i7}, \\ &x_{i11}. \end{aligned}$$

Term x_{i1} :

$$x_{i1} = \frac{1}{\sqrt[4]{T}}\sum_{t=T_{i-1}+1}^{T_i} l_t^{(1)}\left(\xi_t\right).$$

$$\Delta_i = 0.$$

Term x_{i20} :

$$\Delta_i = E\left[x_{i20}|\mathcal{H}_{T_{i-1},T}\right] = 0.$$

Term $x_{i3} + x_{i1}x_{i20}$:

$$\begin{aligned} x_{i3} + x_{i1}x_{i20} &= -\frac{1}{2\sqrt[4]{T^3}}\sum_t l_t^{(1)} m_{2,t} + \frac{1}{2\sqrt[4]{T^3}}\sum_{t,s} l_t^{(1)}\left(m_{2,s} - E\left(m_{2,s}|\mathcal{H}_{T_{i-1},T}\right)\right) \\ &= \frac{1}{2\sqrt[4]{T^3}}\sum_{t\neq s} l_t^{(1)}\left(m_{2,s} - E\left(m_{2,s}|\mathcal{H}_{T_{i-1},T}\right)\right) - \frac{1}{2\sqrt[4]{T^3}}\sum_t l_t^{(1)} E\left(m_{2,s}|\mathcal{H}_{T_{i-1},T}\right) \end{aligned}$$

Δ_i is a martingale.
Terms x_{i5} and x_{i7} :
Δ_i is again a martingale.
Term x_{i11} : $\Delta_i = 0$.

2) Terms in α_t :
We have the terms $x_{i1}, x_{i2}, x_{i3}, x_{i5}, x_{i7}, x_{i11}$.
Term x_{i1} :

$$\begin{aligned}\Delta_i &= E\left(\frac{1}{\sqrt[4]{T}}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}\left(\alpha_t\right)|\mathcal{H}_{T_{i-1},T}\right) \\ &= \frac{1}{\sqrt[4]{T}}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)} E\left(\alpha_t|\mathcal{H}_{T_{i-1},T}\right) \\ &= \frac{1}{\sqrt[4]{T}}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)} E\left(\eta_t|\mathcal{H}_{T_{i-1},T}\right)\end{aligned}$$

because $\alpha_t = E\left(\eta_t|\mathcal{H}_{T_i,T}\right)$

$$\begin{aligned}\|\Delta_i\| &\leq \frac{1}{\sqrt[4]{T}}\sum_{t=T_i+1}^{T_{i+1}} \left\|l_t^{(1)}\right\| \left\|E\left(\eta_t|\mathcal{H}_{T_{i-1},T}\right)\right\| \\ E\|\Delta_i\| &\leq \frac{1}{\sqrt[4]{T}}\sum_{t=T_i+1}^{T_{i+1}} \left(\sup E\left\|l_t^{(1)}\right\|\right)\lambda^{t-T_{i-1}} \\ \sum_{i=1}^{B_N} E\|\Delta_i\| &\leq \frac{const}{\sqrt[4]{T}}\sum_{i=1}^{B_N}\lambda^{B_L} \\ &\leq \frac{const}{\sqrt[4]{T}}\frac{T}{B_L}\lambda^{B_L} \\ &= constT^{3/4}\frac{B_L^{-k}}{B_L}\end{aligned}$$

for any k. Hence $\sum_{i=1}^{B_N} E\|\Delta_i\| \to 0$.
The remaining terms can be treated similarly.

3) Mixed terms:

(R1) We have

$$\begin{aligned}
\Delta_i &= \frac{1}{\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} E\left[l_t^{(2)}\left(\alpha_t, \xi_t\right) | \mathcal{H}_{T_{i-1},T}\right] \\
&= \frac{1}{\sqrt{T}} \sum_{t=T_{i-1}+1}^{T_i} l_t^{(2)}\left(\alpha_t \otimes E\left(\xi_t | \mathcal{H}_{T_{i-1},T}\right)\right) \\
&= 0
\end{aligned}$$

because $E\left(\xi_t | \mathcal{H}_{T_{i-1},T}\right) = 0$. Hence, Lemma 7.2 applies.

Similarly, for (R3), (R5) and (R7), $\Delta_i = 0$. (R2) is a martingale and Lemma 7.3 applies.

For (R9), we can use the fact that $E\left(\xi_t \otimes \xi_t\right)$ is constant and $E\left(\alpha_t | \mathcal{H}_{T_{i-1},T}\right)$ decays exponentially. Indeed we have

$$\begin{aligned}
\Delta_i &= -\frac{1}{2\sqrt[4]{T^3}} \sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)} E(\alpha_t | \mathcal{H}_{T_{i-1},T}) m_{2,t} E\left[(\xi_t \otimes \xi_t)\right] \\
\|\Delta_i\| &\leq \frac{1}{2\sqrt[4]{T^3}} \sum_{t=T_i+1}^{T_{i+1}} \left\| l_t^{(1)} \right\| \|m_{2,t}\| \left\| E(\eta_t | \mathcal{H}_{T_{i-1},T}) \right\| \\
E\|\Delta_i\| &\leq const \frac{1}{\sqrt[4]{T^3}} \sum_{t=T_i+1}^{T_{i+1}} \lambda^{t-T_{i-1}} \leq \frac{\lambda^{B_L}}{\sqrt[4]{T^3}}.
\end{aligned}$$

Hence the conditions of Lemma 7.2 are satisfied.

Yet, terms (R4), (R6) and (R8) remain and will be taken care of later.

For products of mixed terms such that $\sum \alpha_i \geq 1$, it is easy to check that (7.4) and (7.5) are satisfied, since there is α involved.

4) Cross-products involving α_t and ξ_t :

Since the product has α_t involved, as far as $\sum \alpha_i \geq 1$, conditions (7.4) and (7.5) are satisfied. So we only need to concentrate on those terms with $\sum \alpha_i < 1$. They are, $x_{i1}(\xi_t) \cdot x_{i1}(\alpha_t), x_{i1}(\xi_t) \cdot x_{i20}(\alpha_t), x_{i1}(\xi_t) \cdot x_{i11}(\alpha_t), x_{i20}(\xi_t) \cdot x_{i1}(\alpha_t), x_{i1}(\alpha_t) \cdot R1, x_{i11}(\xi_t) \cdot x_{i1}(\alpha_t)$ and $x_{i1}(\xi_t) \cdot R1$. Δ_i is martingale in the first five cases. Hence we can apply Lemma 7.3. We treat the first case in details and omit the other cases.

Term $x_{i1}(\xi_t) \cdot x_{i1}(\alpha_t)$:

$$x_{i1}(\alpha) x_{i1}(\xi) = \frac{1}{\sqrt{T}} \sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}(\alpha_t) \sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}(\xi_s).$$

The associated Δ_i is a martingale. We can apply Lemma 7.3. Remark that

$$\begin{aligned} E\left(\alpha_t \otimes \xi_s | \mathcal{H}_{T_{i-1},T}\right) &= E\left[E\left(\alpha_t | \mathcal{H}_{s,T}\right) \otimes \xi_s | \mathcal{H}_{T_{i-1},T}\right] \\ &= E\left[E\left(\eta_t | \mathcal{H}_{s,T}\right) \otimes \xi_s | \mathcal{H}_{T_{i-1},T}\right], \\ \left\| E\left(\alpha_t \otimes \xi_s | \mathcal{H}_{T_{i-1},T}\right)\right\| &\leq E\left[\left\| E\left(\eta_t | \mathcal{H}_{s,T}\right)\right\| \left\|\xi_s\right\| | \mathcal{H}_{T_{i-1},T}\right] \\ &\leq const \lambda^{t-s} g\left(\eta_{T_{i-1}}, ...\right) \end{aligned}$$

using $\|\xi_s\| \leq 2M$. Hence we have

$$\begin{aligned} & \left| E\left[l_t^{(1)}\left(\alpha_t\right) \sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}\left(\xi_s\right) | \mathcal{H}_{T_{i-1},T} \right] \right| \\ \leq & \sum_{s=T_{i-1}+1}^{T_i} \left| l_t^{(1)} l_s^{(1)} \right| \left| E\left(\alpha_t \otimes \xi_s | \mathcal{H}_{T_{i-1},T}\right) \right| \\ \leq & \ const \sum_{s=T_{i-1}+1}^{T_i} \left| l_t^{(1)} l_s^{(1)} \right| \lambda^{t-s} g\left(\eta_{T_{i-1}}, ...\right). \end{aligned}$$

And

$$\begin{aligned} E\left(\Delta_i^2\right) &\leq \frac{1}{T} \sum_{t=T_i+1}^{T_{i+1}} E\left\{ E\left[l_t^{(1)}\left(\alpha_t\right) \sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}\left(\xi_s\right) | \mathcal{H}_{T_{i-1},T} \right]^2 \right\} \\ &\leq \frac{const}{T} \sum_{t=T_i+1}^{T_{i+1}} \sum_{s=T_{i-1}+1}^{T_i} \lambda^{2(t-s)} E\left[\left(l_t^{(1)} l_s^{(1)} \right)^2 \right] \\ &\leq \frac{const}{T} \left(\sup E \left\| l_t^{(1)} \right\|^4 \right)^2 \sum_{t=T_i+1}^{T_{i+1}} \sum_{s'=0}^{T_i - T_{i-1}} \lambda^{2(s'+t-T_i)} \\ &\leq \frac{const}{T} \left(\frac{1}{1-\lambda^2} \right)^2. \end{aligned}$$

Therefore

$$\sum_i E\left(\Delta_i^2\right) \to 0.$$

Now we turn our attention to the terms that are not martingales. Consider $x_{i11}(\xi_t) \cdot x_{i1}(\alpha_t)$.

$$x_{i11} = \frac{1}{2\sqrt{T}} \sum_{t=T_i+1}^{T_{i+1}} \left[E\left(m_{2,t} | \mathcal{H}_{T_i,T}\right) - E\left(m_{2,t} | \mathcal{H}_{0,T}\right) \right]$$

$$\begin{aligned} & x_{i11}(\xi_t) \cdot x_{i1}(\alpha_t) \\ = & \ \frac{1}{2\sqrt[4]{T^3}} \sum_{t=T_i+1}^{T_{i+1}} m_{2,t} \left[E\left(\xi_t \otimes \xi_t | \mathcal{H}_{T_i,T}\right) - E\left(\xi_t \otimes \xi_t\right) \right] \sum_{s=T_i+1}^{T_{i+1}} l_s^{(1)}(\alpha_s) \end{aligned}$$

For $t = s$, this term cancels out with (R8). For $t \neq s$, we have a martingale and we can apply Lemma 7.3.

Then consider $x_{i1}(\xi_t) \cdot R1$. Using $l_t^{(2)}(\alpha_t, \xi_t) = l_t^{(2)}(\xi_t, \alpha_t)$, this term equals

$$\frac{1}{\sqrt[4]{T^3}} \sum_{t=T_i+1}^{T_{i+1}} l_t^{(2)}(\alpha_t, \xi_t) \sum_{s=T_i+1}^{T_{i+1}} l_s^{(1)}(\xi_s)$$

For $s > t$, it is a martingale. And we know it causes no trouble as we can apply Lemma 7.3. So we consider

$$\begin{aligned} & \frac{1}{\sqrt[4]{T^3}} \sum_t l_t^{(2)}(\alpha_t, \xi_t) \sum_{s \leq t} l_s^{(1)}(\xi_s) \\ = \; & \frac{1}{\sqrt[4]{T^3}} \sum_t l_t^{(2)}(\alpha_t, \xi_t) l_t^{(1)}(\xi_t) && (7.21) \\ & + \frac{1}{\sqrt[4]{T^3}} \sum_t l_t^{(2)}(\alpha_t, \xi_t) \sum_{s<t} l_s^{(1)}(\xi_s) && (7.22) \end{aligned}$$

(7.21) cancels out with (R4). (7.22) can be rewritten as

$$\begin{aligned} & \frac{1}{\sqrt[4]{T^3}} \sum_t l_t^{(2)}(\alpha_t, \xi_t) L_{t-1}(\xi) \\ = \; & \frac{1}{\sqrt[4]{T^3}} \sum_t \left(l_t^{(2)}(\alpha_t, \xi_t) + l_t^{(1)}(\alpha_t) \otimes l_t^{(1)}(\xi_t) \right) L_{t-1}(\xi) && (7.23) \\ & - \frac{1}{\sqrt[4]{T^3}} \sum_t l_t^{(1)}(\alpha_t) \otimes l_t^{(1)}(\xi_t) L_{t-1}(\xi) && (7.24) \end{aligned}$$

(7.23) is again a martingale which causes no problem. Finally, (7.24) cancels out with (R6).

Proof of Corollary 3.2.

By Lemma 4.5 in van der Vaart (1998), contiguity holds if $\ell_T^{\beta}(\theta_T) = dP_{\theta_T,\beta}/dP_{\theta_T} \xrightarrow{d} U$ under P_{θ_T} with $E(U) = 1$. From Theorem 3.1, we have

$$\frac{dP_{\theta_T,\beta}}{dP_{\theta_T}} / \exp\left(\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta_T) - \frac{1}{8}E\left(\mu_{2,t}(\beta,\theta_T)^2\right)\right) \xrightarrow{P} 1$$

under P_{θ_T}. From the CLT for m.d.s, it follows that

$$\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta_T) \xrightarrow{d} N(\beta)$$

under P_{θ_T} where $N(\beta)$ is a Gaussian process with mean 0 and variance $E\left(\mu_{2,t}(\beta,\theta_T)^2\right)/4 \equiv c(\beta,\beta)/4$. Using the expression of the moment generating function of a normal distribution, we have

$$\begin{aligned} E[N(\beta)] &= \exp\left(\frac{c(\beta,\beta)}{8}\right)\exp\left(-\frac{c(\beta,\beta)}{8}\right) \\ &= 1. \end{aligned}$$

Proof of Theorem 3.9 and Lemma 3.10

We have to analyze the difference between

$$Z_T(\beta,\theta_T) = \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta_T) - \frac{1}{8}E\left(\mu_{2,t}(\beta,\theta_T)^2\right) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}d'l_t^{(1)}(\theta_T) + \frac{1}{2}E\left(\left(d'l_t^{(1)}(\theta_T)\right)^2\right) \tag{7.25}$$

where

$$\theta_T = \theta + d/\sqrt{T} \tag{7.26}$$

and d is chosen according to (3.30), and

$$TS_T(\beta,\widehat{\theta}) = \frac{1}{2\sqrt{T}}\sum \mu_{2,t}\left(\beta,\widehat{\theta}\right) - \frac{1}{2T}\widehat{\varepsilon}(\beta)'\widehat{\varepsilon}(\beta), \tag{7.27}$$

where $\widehat{\varepsilon}(\beta)$ is the residual from the OLS regression of $\frac{1}{2}\mu_{2,t}\left(\beta,\hat{\theta}\right)$ on $l_t^{(1)}\left(\hat{\theta}\right)$.

In the theorem, we are only interested in integrals with respect to the measure J. Moreover, this measure has compact support. Hence we can assume that the variable β is restricted to a compact set.

We can easily see that $-\frac{1}{8}E\left(\mu_{2,t}(\beta,\theta_T)^2\right) + \frac{1}{2}E\left(\left(d'l_t^{(1)}(\theta_T)\right)^2\right)$ are continuous functions of θ, converging uniformly in β to

$$-\frac{1}{8}E\left(\mu_{2,t}(\beta,\theta_0)^2\right) + \frac{1}{2}E\left(\left(d'l_t^{(1)}(\theta_0)\right)^2\right). \tag{7.28}$$

Let

$$\hat{d} = \hat{d}(\beta) = \left(\frac{1}{T}\sum_{t=1}^{T} l_t^{(1)}\left(\hat{\theta}\right) \otimes l_t^{(1)}\left(\hat{\theta}\right)\right)^{-1} \left(\frac{1}{2T}\sum_{t=1}^{T} \mu_{2,t}\left(\hat{\theta},\beta\right) l_t^{(1)}\left(\hat{\theta}\right)\right).$$

Denote $y_t = \frac{1}{2}\mu_{2,t}\left(\hat{\theta}\right)$, $x_t = l_t^{(1)}\left(\hat{\theta}\right)$, $y = (y_1, ..., y_T)'$ and $X = (x_1, ..., x_T)'$. Using these notations, $\hat{d} = (X'X)^{-1} X'y$ and

$$\begin{aligned} & \frac{1}{4T}\sum_t \left[\mu_{2,t}\left(\hat{\theta},\beta\right)\right]^2 - \hat{d}'\hat{I}\left(\hat{\theta}\right)\hat{d}/T \\ = & \left(y'y - y'X(X'X)^{-1}X'y\right)/T \\ = & y'\left[I - X(X'X)^{-1}X'\right]y/T \\ = & y'M_X M_X y/T \\ = & \widehat{\varepsilon(\beta)}'\widehat{\varepsilon(\beta)}/T \end{aligned}$$

where $M_X = I - X(X'X)^{-1}X'$ is idempotent. Obviously our assumptions guarantee the consistency of the ML estimator. Then it is now an elementary exercise to show that

$$\hat{d}(\beta) \to d(\beta) \tag{7.29}$$

and consequently

$$\begin{aligned} \frac{1}{2T}\widehat{\varepsilon(\beta)}'\widehat{\varepsilon(\beta)} & \to \frac{1}{8}E\mu_{2,t}(\theta_0,\beta)^2 - \frac{1}{2}d'I(\theta_0)d & (7.30) \\ & = \frac{1}{2}E\left[\left(\frac{\mu_{2,t}(\theta_0,\beta)}{2} - d'l_t^{(1)}(\theta_0)\right)^2\right] & (7.31) \end{aligned}$$

by (3.30). Hence it is sufficient for us to show that

$$\begin{aligned} & \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta_T) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d'l_t^{(1)}(\theta_T) - \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}\left(\beta,\widehat{\theta}\right) \\ = & \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta_T) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d'l_t^{(1)}(\theta_T) - \left(\frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}\left(\beta,\widehat{\theta}\right) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d'l_t^{(1)}\left(\widehat{\theta}\right)\right) \end{aligned}$$

converges (uniformly in β) to 0. So define the function

$$Y_T(\beta,\theta) = \frac{1}{2\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d'l_t^{(1)}(\theta). \tag{7.32}$$

Observe that our conditions guarantee that the ML estimator is $\sqrt{T}$ consistent. Hence it is sufficient to show that for all M

$$\sup_{\beta,\|\theta-\theta_0\|\le M/\sqrt{T}} |Y_T(\beta,\theta) - Y_T(\beta,\theta_0)| \to 0 \tag{7.33}$$

Obviously Y_T is at least twice continuously differentiable as a function of θ, and we can easily see that its second derivative is $O(\sqrt{T})$. Hence to show (7.33) it is sufficient to show that the first derivative is $o(\sqrt{T})$ or equivalently

$$\frac{\partial}{\partial\theta}\left(\frac{1}{2T}\sum_{t=1}^{T}\mu_{2,t}(\beta,\theta)-\frac{1}{T}\sum_{t=1}^{T}d'l_t^{(1)}(\theta)\right)\to 0 \tag{7.34}$$

Remark 1. *Here we will use "conventional" calculus for partial derivatives, because the direct evaluation of the terms appearing in this proof is relatively easy.*

Remark 2. *Since the second derivative is $O(\sqrt{T})$, and the range of the arguments is $O(1/\sqrt{T})$, the changes in the first derivative are $O(1)$. Hence it is sufficient to show the relationship (7.34) only for one value of θ.*

Moreover, it is easily seen that these results prove the first part of Lemma 3.10. For the second part, the CLT, we apply the proposition of Andrews (1994, page 2251). The finite dimensional convergence follows from the fact that $\mu_{2,t}(\beta,\theta_0)$ is a martingale difference sequence and from the moment conditions imposed in Assumption 4, so that the CLT for m.d.s. applies. The proof of stochastic equicontinuity can be done along the line of Andrews and Ploberger (1996, Proof of Theorem 1).

Let us first state a lemma. Its proof will be given after the proof of the theorem. To simplify our notation: **All of the subsequent statements about convergence should be understood as uniform convergence in β.**

Lemma 7.8. *We have*

$$\frac{1}{T}\sum_t\frac{\partial\mu_{2,t}}{\partial\theta}=-\frac{1}{T}\sum_t\mu_{2,t}\frac{\partial l_t}{\partial\theta}+o_P(1)$$

To establish (7.34), we have to show that

$$\frac{1}{2T}\sum_t\frac{\partial\mu_{2,t}}{\partial\theta}-\frac{1}{T}\sum_{t=1}^{T}d'l_t^{(2)}(\theta)\xrightarrow{P}0. \tag{7.35}$$

The average of the second derivatives equals the negative Information matrix,

$$\frac{1}{T}\sum_{t=1}^{T}l_t^{(2)}(\theta)\xrightarrow{P}-I(\theta) \tag{7.36}$$

and from Lemma 7.8, it follows that

$$\frac{1}{T}\sum_t\frac{\partial\mu_{2,t}}{\partial\theta}\xrightarrow{P}-cov\left(\mu_{2,t},\frac{\partial l_t}{\partial\theta}\right). \tag{7.37}$$

Then (7.34) is an easy consequence of the definition of d in (3.30).

We now have shown the first part of the theorem. It remains to prove the second part of the theorem. Essentially we are establishing some kind of pivotal property of our test statistic. $TS_T\left(\widehat{\theta},\beta\right)$ is a function of the data alone, so its distribution is determined by the underlying distribution of the data. We did establish that the process $TS_T\left(\widehat{\theta},\beta\right)$ converges in distribution, hence its probability distributions remain uniformly tight. For every $\varepsilon > 0$ we can find compact sets of continuous functions so that their probabilities are at least $1-\varepsilon$. The Arzela-Ascoli theorem characterizes the elements of compact sets to be equicontinuous. Equicontinuity implies that we can approximate the integrals $\int \exp(TS_T(\beta,\widehat{\theta_T}))d\nu(\beta,d)$ by finite sums $\sum \nu_i \exp(TS_T(\beta_i,\widehat{\theta_T}))$. Hence it is sufficient to show that the distributions of the finite-dimensional vectors $\left(TS_T(\beta_i,\widehat{\theta_T}) : 1 \leq i \leq N\right)$ are asymptotically the same for all θ such that $\|\theta-\theta_0\| \leq M/\sqrt{T}$ for M arbitrary. Asymptotically, the density between probabilities corresponding to parameters $\theta_0+h/\sqrt{T}$, $\theta_0+k/\sqrt{T}$ is lognormal with mean $O(\|h-k\|)$ and variance $O(\|h-k\|^2)$. Hence, for every $\varepsilon > 0$ we can find finitely many parameter values, say $h_1,..h_j$ so that for every h with $\|h\| \leq M$, there is an h_i such that the total variation of the difference of the probability distributions corresponding to parameters $\theta_0+h/\sqrt{T}$ and $\theta_0+h_i/\sqrt{T}$ is smaller than ε. Hence it is sufficient to show that the distributions of $\left(TS_T(\beta_i,\widehat{\theta_T}) : 1 \leq i \leq N\right)$ are the same when the data are generated by $\theta_0+h_i/\sqrt{T}$. To show this, we can apply Lemma 3.10. Under P_{θ_0}, the $TS_T(\beta_i,\widehat{\theta_T})$ are normalized sums of martingale-differences (plus constants), and elementary calculations show that

$$\log\frac{dP_{\theta_0+h_i/\sqrt{T}}}{dP_{\theta_0}} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} h_i' l_t^{(1)}(\theta_T) + \frac{1}{2}E\left(h_i' l_t^{(1)}(\theta_T)\right)^2 \to 0. \tag{7.38}$$

Hence it can (from the multivariate CLT) easily be seen that the **joint distribution** of $TS_T(\beta_i,\widehat{\theta_T})$ and the logarithm of the densities is a multivariate normal distributions. It is easily verifiable that our construction of the $TS_T(\beta_i,\widehat{\theta_T})$ implies that asymptotically it is uncorrelated and hence independent from the logarithm of the densities. Our proposition is then an easy consequence of this fact.

So it remains to show the lemma:

Proof of Lemma 7.8. First we are rewriting $\frac{\partial \mu_{2,t}}{\partial\theta_k}$. Here we omit the argument $E(\eta_t \otimes \eta_s)$.

$$\mu_{2,t} = l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)} + 2\sum_{s>0} l_t^{(1)} \otimes l_{t-s}^{(1)}$$

$$\frac{\partial}{\partial\theta_k}(l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)}) = \frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j}\right) = \frac{\partial^3 l_t}{\partial\theta_k\partial\theta_i\partial\theta_j} + \frac{\partial^2 l_t}{\partial\theta_k\partial\theta_i}\frac{\partial l_t}{\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial^2 l_t}{\partial\theta_k\partial\theta_j}$$

¿From the third Bartlett identity,

$$m_{3,t} = \frac{\partial^3 l_t}{\partial\theta_k\partial\theta_i\partial\theta_j} + \frac{\partial l_t}{\partial\theta_j}\frac{\partial^2 l_t}{\partial\theta_k\partial\theta_i} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial^2 l_t}{\partial\theta_k\partial\theta_j} + \frac{\partial l_t}{\partial\theta_k}\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j}\frac{\partial l_t}{\partial\theta_k}$$

is a martingale difference sequence and therefore $\frac{1}{T}\sum_{t=1}^{T} m_{3,t} = o_p(1)$.

$$\begin{aligned}
\frac{\partial}{\partial\theta_k}\frac{1}{T}\sum_{t=1}^{T}(l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)}) &= \frac{1}{T}\sum_{t=1}^{T} m_{3,t} - \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j}\right]\frac{\partial l_t}{\partial\theta_k} \\
&= o_p(1) - \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j}\right]\frac{\partial l_t}{\partial\theta_k}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta_k}\frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\frac{\partial l_t}{\partial\theta_i}\frac{\partial l_{t-s}}{\partial\theta_j} &= \frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\left[\frac{\partial^2 l_t}{\partial\theta_k\partial\theta_i}\frac{\partial l_{t-s}}{\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial^2 l_{t-s}}{\partial\theta_k\partial\theta_j}\right] \\
&= \frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\left[\frac{\partial^2 l_t}{\partial\theta_k\partial\theta_i} + \frac{\partial l_t}{\partial\theta_k}\frac{\partial l_t}{\partial\theta_i}\right]\frac{\partial l_{t-s}}{\partial\theta_j} \\
&\quad + \frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\frac{\partial l_t}{\partial\theta_i}\frac{\partial^2 l_{t-s}}{\partial\theta_k\partial\theta_j} \\
&\quad - \frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j}\frac{\partial l_{t-s}}{\partial\theta_k} \\
&= o_p(1) - \frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta j}\frac{\partial l_{t-s}}{\partial\theta_k}
\end{aligned}$$

because $\frac{\partial^2 l_t}{\partial\theta_k\partial\theta_i} + \frac{\partial l_t}{\partial\theta_k}\frac{\partial l_t}{\partial\theta i}$ and $\frac{\partial l_t}{\partial\theta_i}$ are m.d.s. Therefore, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\mu_{2,t}}{\partial\theta_k} &= -\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j} + \frac{2}{T}\sum_{t=1}^{T}\sum_{s>0}\frac{\partial l_t}{\partial\theta_i}\frac{\partial l_{t-s}}{\partial\theta j}\right]\frac{\partial l_t}{\partial\theta_k} + o_P(1) \\
&= -\widehat{cov}\left(\mu_{2,t}, \frac{\partial l_t}{\partial\theta_k}\right) + o_P(1)
\end{aligned}$$

where $\widehat{cov}$ denotes the empirical covariance. It is now an easy exercise to show that $\widehat{cov}\left(\mu_{2,t}, \frac{\partial l_t}{\partial\theta_k}\right) \to cov\left(\mu_{2,t}, \frac{\partial l_t}{\partial\theta_k}\right)$.

Proof of Proposition 5.1.

We do the proof for the case where S_t takes two values only. The generalization to three regimes is immediate. We use the following notation $z_t = \ln(P_t)$ and $w_t = \ln(D_t)$ and we reparametrize slightly (5.3) so that

$$\begin{aligned}
z_t &= a_0 + a_1 w_t + y_t \\
y_t &= \alpha_{s_t} + \sum_{j=1}^{l}\gamma_{s_t,j}y_{t-j} + \varepsilon_t.
\end{aligned}$$

In the two-step approach, the parameters are such that

$$\sum \hat{y}_t = 0, \tag{7.39}$$

$$\sum w_t \hat{y}_t = 0, \tag{7.40}$$

$$\sum \hat{\varepsilon}_t^i P\left(S_t = i | \hat{y}_{t-1}, ..., \hat{y}_1\right) = 0, \tag{7.41}$$

$$\sum \hat{y}_{t-j} \hat{\varepsilon}_t^i P\left(S_t = i | \hat{y}_{t-1}, ..., \hat{y}_1\right) = 0, \;\; j = 1, ..., l, \; i = 0, 1. \tag{7.42}$$

The last two equations are obtained using the expression of the score given by Hamilton (1994, page 692) and the notation

$$\begin{aligned} \hat{\varepsilon}_t^i &= \hat{y}_t - \hat{\alpha}_i - \sum_{j=1}^{l} \hat{\gamma}_{i,j} \hat{y}_{t-j}, \\ \hat{y}_t &= z_t - \hat{a}_0 - \hat{a}_1 w_t \\ &= \left(z_t - \overline{z}\right) - \hat{a}_1 \left(w_t - \overline{w}\right) \end{aligned}$$

Note that there is a potential problem of identification as $\sum \hat{y}_t = 0$ by construction. Therefore, we do not estimate a_0 when we do global MLE, instead we demean the time series z_t and w_t. To compute the global MLE, we use the equation

$$\left(1 - \sum_{j=1}^{l} \gamma_{s_t,j} L^j\right) \left(z_t - \overline{z}\right) = a_1 \left(1 - \sum_{j=1}^{l} \gamma_{s_t,j} L^j\right) \left(w_t - \overline{w}\right) + \alpha_{s_t} + \varepsilon_t.$$

Hence the conditional log-likelihood equals

$$\begin{aligned} & \ln f\left(z_t | w_t, z_{t-1}, w_{t-1}, .., z_1, w_1; s_t\right) \\ = & -\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2\sigma^2} \left\{ \left(1 - \sum_{j=1}^{l} \gamma_{s_t,j} L^j\right) \left(\left(z_t - \overline{z}\right) - a_1 \left(w_t - \overline{w}\right)\right) - \alpha_{s_t} \right\}^2 \end{aligned}$$

Using Hamilton (1994), the scores can be written as

$$\frac{\partial L}{\partial \delta} = \sum_t \sum_{s_t = 0,1} \frac{\partial}{\partial \delta} \ln f\left(z_t | w_t, z_{t-1}, w_{t-1}, .., z_1, w_1; s_t\right) P\left(S_t = s_t | z_{t-1}, w_{t-1}, ..., z_1, w_1\right).$$

Hence we have

$$\frac{\partial L}{\partial \alpha_i} = \frac{1}{\sigma^2} \sum_t \hat{\varepsilon}_t^i P\left(S_t = i | z_{t-1}, w_{t-1}, ..., z_1, w_1\right) = 0, \;\; i = 0, 1 \tag{7.43}$$

$$\frac{\partial L}{\partial \gamma_{i,j}} = \frac{1}{\sigma^2} \sum_t \hat{y}_{t-j} \hat{\varepsilon}_t^i P\left(S_t = i | z_{t-1}, w_{t-1}, ..., z_1, w_1\right) = 0, \;\; j = 1, ..., l, \;\; i = 0, 1 \tag{7.44}$$

As the relevant information (for S_t) contained in $\sigma\left(z_{t-1}, w_{t-1}, ..., z_1, w_1\right)$ is the same as that contained in $\sigma\left(\hat{y}_{t-1}, ..., \hat{y}_1\right)$, (7.43) and (7.44) coincide with (7.41) and (7.42).

$$\frac{\partial L}{\partial a_1} = \frac{1}{\sigma^2}\sum_t \sum_{i=0,1}\left((w_t-\overline{w})-\sum_j \hat{\gamma}_{i,j}(w_{t-j}-\overline{w})\right)\hat{\varepsilon}_t^i P(S_t=i|z_{t-1},w_{t-1},...,z_1,w_1)=0. \tag{7.45}$$

Note that $\hat{\gamma}_{i,j}$ is selected so that (7.43) and (7.44) hold. (7.45) will be guaranteed if

$$\sum_t \sum_{i=0,1}(w_t-\overline{w})\left(\hat{y}_t-\hat{\alpha}_i-\sum_{j=1}^{l}\hat{\gamma}_{i,j}\hat{y}_{t-j}\right)P(S_t=i|z_{t-1},w_{t-1},...,z_1,w_1) = 0 \tag{7.46}$$

$$\sum_t \sum_{i=0,1}(w_{t-j}-\overline{w})\hat{\varepsilon}_t^i P(S_t=i|z_{t-1},w_{t-1},...,z_1,w_1) = 0$$

(7.46) holds if

$$\sum_t (w_t-\overline{w})\hat{y}_t = 0 \tag{7.47}$$

$$\sum_t \sum_{i=0,1}(w_t-\overline{w})\left(\hat{\alpha}_i+\sum_{j=1}^{l}\hat{\gamma}_{i,j}\hat{y}_{t-j}\right)P(S_t=i|z_{t-1},w_{t-1},...,z_1,w_1) = 0 \tag{7.48}$$

$j,k=1,...,l$ where $\bar{y}=\sum_t \hat{y}_t/T$.

$$\begin{aligned}(7.47) \;&\Leftrightarrow\; \sum_t w_t(\hat{y}_t-\bar{y})=0\\ &\Leftrightarrow\; \sum_t w_t\left((z_t-\overline{z})-\hat{a}_1(w_t-\overline{w})\right)=0,\end{aligned}$$

corresponds to (7.40). The other equations overidentify the parameters but are satisfied in large sample as long as w_t is strictly exogenous. So far, we have shown that the two-step estimators coincide asymptotically with the global MLE. Now we turn our attention toward the independence.

To show the independence, we need to show that the Hessian is block diagonal. We consider the Hessian for the true values of the parameters.

$$\frac{\partial^2 L}{\partial a_1 \partial \alpha_i} = -\frac{1}{\sigma^2}\sum_t\left((w_t-\overline{w})-\sum_j \gamma_{i,j}(w_{t-j}-\overline{w})\right)P(S_t=i|z_{t-1},w_{t-1},...,z_1,w_1)$$

$$E\left[\frac{\partial^2 L}{\partial a_1 \partial \alpha_i}\right]=0$$

because

$$\begin{aligned}&E\left[(w_{t-j}-\overline{w})P(S_t=1|z_{t-1},w_{t-1},...,z_1,w_1)\right]\\ =\;& E\left[(w_{t-j}-\overline{w})P(S_t=1|y_{t-1},...,y_1)\right]\\ =\;& E\left[(w_{t-j}-\overline{w})S_t\right]\\ =\;& E(w_{t-j}-\overline{w})E(S_t)\\ =\;& 0,\; j=0,1,...,l,\end{aligned}$$

assuming that w_t is uncorrelated with $y_t, ..., y_T$.

$$\begin{aligned}
\frac{\partial^2 L}{\partial a_1 \partial \gamma_{i,j}} &= -\frac{1}{\sigma^2} \sum_t \left((w_t - \overline{w}) - \sum_k \gamma_{i,k} (w_{t-k} - \overline{w}) \right) y_{t-j} P(S_t = i | z_{t-1}, w_{t-1}, ..., z_1, w_1) \\
&\quad -\frac{1}{\sigma^2} \sum_t (w_{t-j} - \overline{w}) \varepsilon_t^i P(S_t = i | z_{t-1}, w_{t-1}, ..., z_1, w_1) \\
E\left[\frac{\partial^2 L}{\partial a_1 \partial \gamma_{i,j}} \right] &= 0.
\end{aligned}$$

In conclusion, $\hat{a}_1$ is independent of $\left(\hat{\alpha}_i, \hat{\gamma}_{i,j}\right)$ if z_t is strictly exogenous.

8. References

Andrews, D.W.K, (1993). "Tests for Parameter Instability and Structural Change Point", *Econometrica* 61, 821-856.

Andrews, D. (1994) "Empirical process methods in econometrics", in *Handbook of Econometrics*, Vol. IV, ed. by R.F. Engle and D.L. McFadden.

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

Andrews, D.W.K., Ploberger, W. (1995). "Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative", *The Annals of Statistics* 23, 1609-1629.

Andrews, D. and W. Ploberger (1996) "Testing for Serial Correlation Against an ARMA(1,1) Process", *Journal of the American Statistical Association*, Vol. 91, 1331-1342.

Bartlett, M. S. (1953a) "Approximate confidence intervals", *Biometrika*, 40, 12-19.

Bartlett, M. S. (1953b) "Approximate confidence intervals, II. More than one parameter", *Biometrika*, 40, 306-317.

Bohl, M. (2001) "Periodically Collapsing Bubbles in the US Stock Market", forthcoming in *International Review of Economics and Finance*.

Chesher, A. (1984) "Testing for Neglected Heterogeneity", *Econometrica*, 22, 865-872.

Chirinko, R. and H. Schaller (2001) "Business Fixed Investment and "Bubbles": The Japanese Case", *American Economic Review*, 91, 663-680.

Cho, J.S. and H. White (2003) "Testing for Regime Switching", mimeo, UCSD.

Cox D.R. and H. D. Miller (1965) *The Theory of Stochastic Processes*, Methuen & Co. Ltd, London, UK.

Davidson, R. and J. MacKinnon (1991) "Une nouvelle forme du test de la matrice d'information", *Annales d'Economie et de Statistique*, N. 20/21, 171-192.

Davies, R.B. (1977) "Hypothesis testing when a nuisance parameter is present only under the alternative" *Biometrika*, **64**, 247-254.

Davies, R.B. (1987) "Hypothesis testing when a nuisance parameter is present only under the alternative" *Biometrika*, **74**, 1, 33-43.

Diba, B. and H. Grossman (1988) "Explosive Rational Bubbles in Stock Prices", *American Economic Review*, 78, 520-530.

Doukhan, P. (1994) *Mixing: Properties and Example*, Springer-Verlag, New York.

Evans, G. (1991) "Pitfalls in Testing for Explosive Bubbles in Asset Prices", *American Economic Review*, 4, 922-930.

Garcia, R., 1998. Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models, International Economic Review 39, 763-788.

Hall, S., Z. Psaradakis, and M. Sola (1999) "Detecting Periodically Collapsing Bubbles: A Markov-Switching Unit Root Test", *Journal of Applied Econometrics*, 14, 143-154.

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. (1992) "The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*, 7, S61-S82.

Lang, S. (1993) *Real and Functional Analysis*, Springer-Verlag, New York.

Lee, L-F. and A. Chesher (1984) "Specification Testing when Score Test Statistics Are Identically Zero", *Journal of Econometrics*, 31, 121-149.

Lucas, R. (1978) "Asset Prices in an Exchange Economy", *Econometrica*, 66, 429-445..

Murray, M.K and J.W. Rice(1993): "Differential Geometry and Statistics", Chapman-Hall, London.

Pagan, A. and K. Sossounov (2003) "A simple framework for analysing bull and bear markets", *Journal of Applied Econometrics*, **18**, 23-46.

Psaradakis, Z., M. Sola, and F. Spagnolo (2001) "A simple procedure for detecting periodically collapsing rational bubbles", *Economics Letters*, 317-323.

Shiller, R. (2000) *Irrational Exuberance*, Princeton University Press, Princeton, New Jersey.

Strasser, H. (1995) *Mathematical Theory of Statistics*. Springer Verlag, New York.

Taylor, M. P. and D. Peel (1998) "Periodically collapsing stock price bubbles: a robust test", *Economics Letters*, 61, 221-228.

van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK.

van Norden, S. and H. Schaller (1993) "The Predictability of Stock Market Regime: Evidence from the Toronto Stock Exchange", *Review of Economics and Statistics*, 75, 505-510.

van Norden, S. and R. Vigfusson (1996) "Avoiding the Pitfalls: Can Regime-Switching Tests Reliably Detect Bubbles", *Studies in Nonlinear Dynamics and Econometrics*, 3, 1-22.

White, H. (1982) "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-25.

Yao, J.-F. and J.-G. Attali (2000) "On stability of nonlinear AR processes with Markov switching", *Advances in Applied Probability*, 32, 394-407.

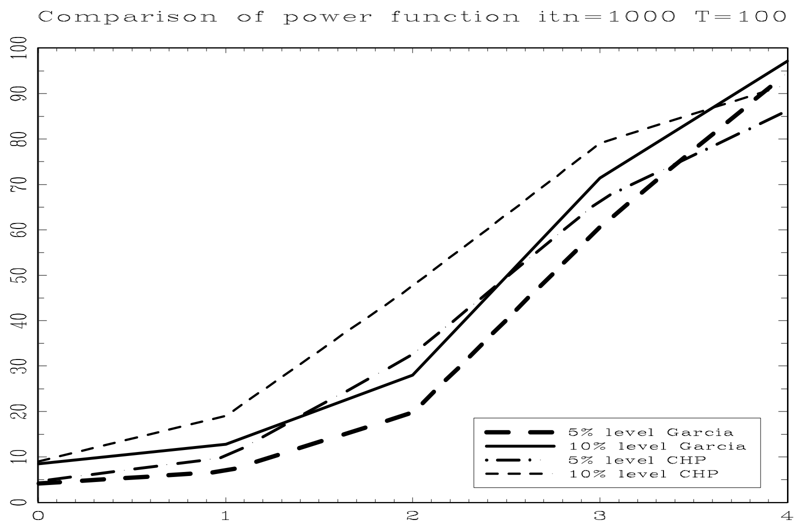

Figure 8.1: Comparison of size-corrected powers

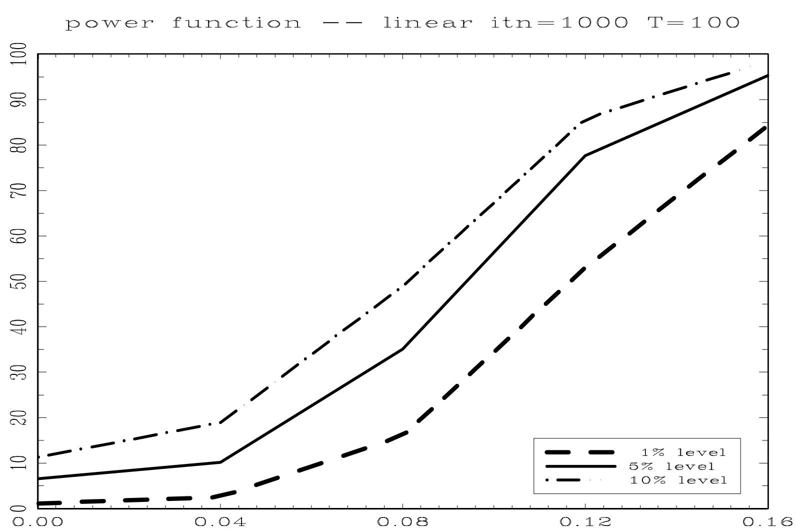

Figure 8.2: Linear model with intercept

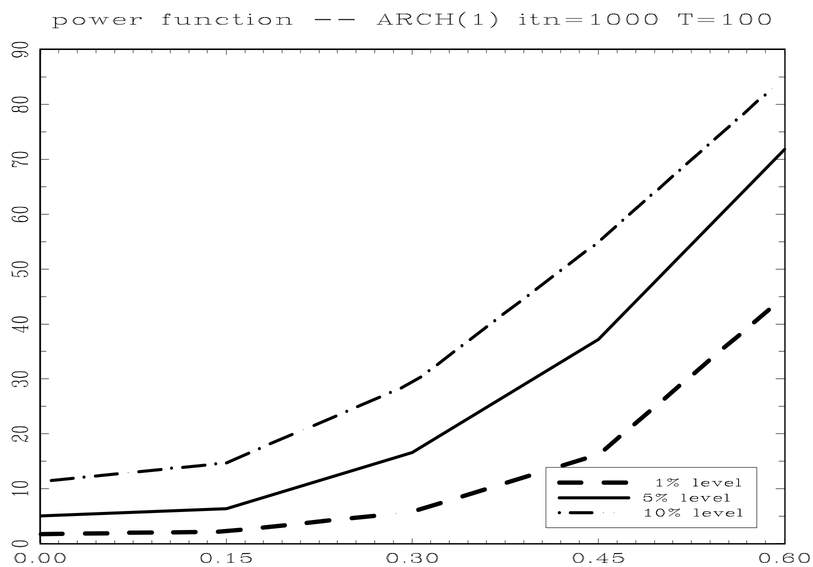

Figure 8.3: ARCH(1)

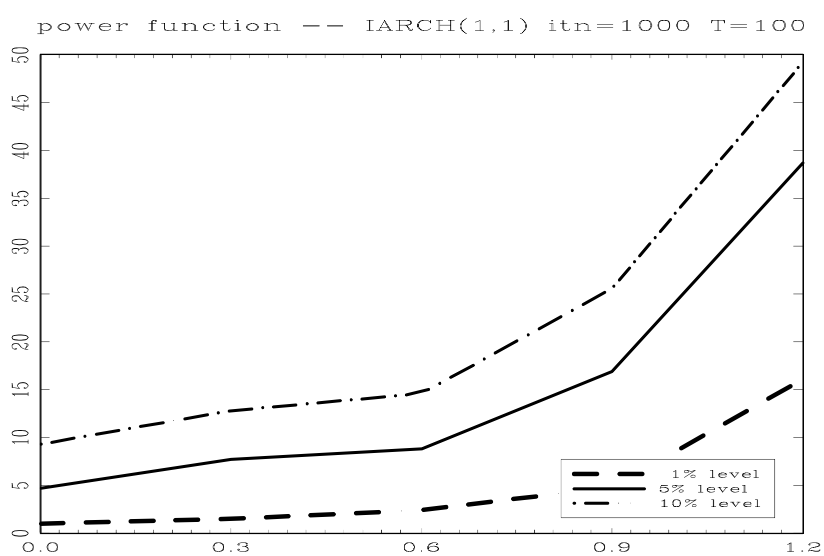

Figure 8.4: IGARCH(1,1)