Econometrica Supplementary Material

SUPPLEMENT TO "OPTIMAL TEST FOR MARKOV SWITCHING PARAMETERS"

(*Econometrica*, Vol. 82, No. 2, March 2014, 765–784)

BY MARINE CARRASCO, LIANG HU, AND WERNER PLOBERGER

THIS SUPPLEMENT IS ORGANIZED AS FOLLOWS. Section A.1 reviews various potential applications of our test. Section A.2 establishes Assumption 2 for ARCH(1) and GARCH(1, 1) models with normal errors. Section A.3 derives the expression and asymptotic distribution of our tests for Hamilton's model. Corresponding critical values are given in Section A.4. Section A.5 gives the estimates of Hamilton's model for U.S. real GNP. Appendix B defines the tensor notations used to derive the fourth-order expansion of the likelihood. These notations are interesting in their own right, as they could be used in other econometric problems involving higher-order expansions. Appendix C collects the proofs of Theorems 3.1, 3.2, and 4.1. Appendix D collects the proofs of the remaining results of Sections 4 and 5. Section D.4 gives Lemma D.9, which establishes necessary and sufficient conditions for expTS to have power in the context of an autoregressive model.

APPENDIX A: EXAMPLES

A.1. *Review of Applications*

Various extensions of Example 3.4 have been applied in macroeconomics and finance.

EXAMPLE A.1—Markov Switching GARCH Model: Markov switching GARCH models are increasingly used in finance; see Hamilton and Susmel (1994), Dueker (1997), Gray (1996), Haas, Mittnik, and Paolella (2004), among others. Consider the model

$$\begin{cases} \varepsilon_t = z_t \sigma_t, \\ \sigma_t^2 = \alpha_0(S_t) + \alpha_1(S_t)\varepsilon_{t-1}^2 + \beta_1(S_t)\sigma_{t-1}^2, \\ z_t \sim \text{i.i.d. } \mathcal{N}(0,1), \end{cases}$$

where S_t is a homogeneous Markov chain with k-dimensional state space. Then, $\vartheta_t = S_t$ is bounded, Markov of order 1, and geometric ergodic provided its transition probabilities belong to $(0, 1)$. The estimation of this model is particularly tedious. This model has been successfully tested in Hu and Shin (2008) using our test procedure.

 DOI: 10.3982/ECTA8609

EXAMPLE A.2—State Space Model: Assume that the dynamic of an observable vector y_t can be described as

$$y_t = A'x_t + H'\xi_t + w_t,$$
$$\xi_{t+1} = F\xi_t + v_{t+1},$$

where v_t and w_t are uncorrelated white noises and x_t is a vector of exogenous or predetermined variables.

The state vector ξ_t is not observable. The state space models are very popular because they can be easily estimated by Kalman filter. To simplify our discussion, assume that ξ_t is scalar. A way to test the null hypothesis that ξ_t is constant $H_0 : \xi_t = \xi_0$ is to test that the variance of v_t (σ^2, say) is equal to zero. Various difficulties arise. First, the parameter of interest (σ^2) is on the boundary of the parameter space under the null. Second, the coefficient F is not identified under H_0. Hence, testing $H_0 : \sigma^2 = 0$ is nonstandard. These issues are addressed in Andrews (1999, 2001).

EXAMPLE A.3—Non-Markov Random Coefficient Model: Consider a stochastic volatility model

$$\begin{cases} y_t = \mu + \theta y_{t-1} + z_t\sigma_t, \\ \sigma_t = \exp(v_t), \\ v_t = \alpha + \beta v_{t-1} + e_t + \delta e_{t-1}, \\ z_t \sim \text{i.i.d. } \mathcal{N}(0, 1), \end{cases}$$

where z_t and e_t are independent and e_t is i.i.d. $(0, \tau^2)$. $\vartheta_t = (v_t, e_t)'$ is Markov, geometric ergodic provided $|\beta| < 1$, and e_t has positive density around 0 although σ_t itself is not Markov. This model is easy to estimate under the null hypothesis where σ_t is constant.

A.2. *Assumption 2 for ARCH and GARCH Models*

To apply our test to a Markov switching GARCH model as described in Example A.1, we need to make sure that Assumption 2 is satisfied for such a model. First, we give a detailed proof that Assumption 2 is satisfied for the ARCH(1) model. Second, we give a sketch of the proof for the GARCH(1, 1) model.

We consider an ARCH(1) model with normal error:

$$\begin{cases} y_t = \sigma_t z_t, \\ \sigma_t^2 = \omega + \alpha y_{t-1}^2, \end{cases}$$

where $\alpha > 0$, $\omega > 0$, and z_t i.i.d. $\mathcal{N}(0,1)$. Let $\theta = (\alpha, \omega)'$ and θ_0 be the true value of the parameter vector. The conditional log-density is given by

$$l_t(\theta) = -\frac{1}{2}\left(\ln \sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right),$$

where $\sigma_t^2(\theta) = \omega + \alpha y_{t-1}^2$. We can compute the kth derivatives of l_t recursively (see, for instance, Jensen and Rahbek (2004a)):

$$\frac{\partial^k l_t(\theta)}{\partial \alpha^k} = \frac{(-1)^k (k-1)!}{2}\left(1 - k\frac{y_t^2}{\sigma_t^2(\theta)}\right)\frac{y_{t-1}^{2k}}{\sigma_t^{2k}(\theta)}, \quad k = 1, 2, \ldots,$$

$$\frac{\partial^k l_t(\theta)}{\partial \omega^k} = \frac{(-1)^k (k-1)!}{2}\left(1 - k\frac{y_t^2}{\sigma_t^2(\theta)}\right)\frac{1}{\sigma_t^{2k}(\theta)}, \quad k = 1, 2, \ldots.$$

Consider a neighborhood of θ_0, denoted by $\mathcal{N}$, such that $0 < \omega_l < \omega < \omega_u$ and $0 < \alpha_l < \alpha < \alpha_u$. Observe that, for $\theta \in \mathcal{N}$, we have

$$\begin{aligned}
\frac{y_t^2}{\sigma_t^2(\theta)} &= \frac{\sigma_t^2(\theta_0) z_t^2}{\sigma_t^2(\theta)} \\
&= \frac{(\omega_0 + \alpha_0 y_{t-1}^2) z_t^2}{\omega + \alpha y_{t-1}^2} \\
&\le \left(\frac{\omega_0}{\omega_l} + \frac{\alpha_0}{\alpha_l}\right) z_t^2, \\
\frac{y_{t-1}^{2k}}{\sigma_t^{2k}(\theta)} &= \frac{y_{t-1}^{2k}}{(\omega + \alpha y_{t-1}^2)^k} \le \left(\frac{1}{\alpha_l}\right)^k, \\
\frac{1}{\sigma_t^{2k}(\theta)} &\le \left(\frac{1}{\omega_l}\right)^k.
\end{aligned}$$

It follows that, for $k = 1, 2, \ldots, 5$,

$$E \sup_{\theta \in \mathcal{N}} \left| \frac{\partial^k l_t(\theta)}{\partial \alpha^k} \right|^{20} < \infty,$$

$$E \sup_{\theta \in \mathcal{N}} \left| \frac{\partial^k l_t(\theta)}{\partial \omega^k} \right|^{20} < \infty.$$

Hence, Assumption 2 holds.

Consider a GARCH(1, 1) model with normal error:

$$\begin{cases} y_t = \sqrt{h_t(\theta)} z_t, \\ h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta), \end{cases}$$

where z_t i.i.d. $\mathcal{N}(0,1)$. Following Jensen and Rahbek (2004b), the initial variance $h_0(\theta)$ is parameterized as $\gamma = h_0(\theta)$. Let $\theta = (\alpha, \beta, \omega, \gamma)'$, where all the parameters are positive. The true value of the parameter vector is denoted θ_0. The conditional log-density is given by

$$l_t(\theta) = \ln h_t(\theta) + \frac{y_t^2}{h_t(\theta)}.$$

The derivative of l_t with respect to θ_i, the ith component of θ, is

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = \left(1 - \frac{y_t^2}{h_t(\theta)}\right)\frac{\partial h_t(\theta)/\partial \theta_i}{h_t(\theta)}.$$

The higher-order derivatives are functions of $y_t^2/h_t(\theta)$ and $\partial^k h_t(\theta)/\partial\theta_i^k/h_t(\theta)$.

Observe that

$$\begin{aligned}\frac{y_t^2}{h_t(\theta)} &= \frac{h_t(\theta_0)}{h_t(\theta)} z_t^2 = \left(\frac{\omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 h_{t-1}(\theta_0)}{\omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta)}\right) z_t^2 \\ &\le \left(\frac{\omega_0}{\omega_l} + \frac{\alpha_0}{\alpha_l} + \frac{\beta_0 h_{t-1}(\theta_0)}{\beta_l h_{t-1}(\theta)}\right) z_t^2.\end{aligned}$$

Replacing recursively $h_{t-1}(\theta_0)/h_{t-1}(\theta)$ by $h_{t-2}(\theta_0)/h_{t-2}(\theta)$, etc., we obtain an upperbound for $y_t^2/h_t(\theta)$ which is a function of θ_0, θ_l, and $z_t^2, z_{t-1}^2, \ldots$. Upperbounds for terms $\partial^k h_t(\theta)/\partial\theta_i^k/h_t(\theta)$ are given in Jensen and Rahbek (2004b); see, for instance, Lemmas 3 and 9 for the derivatives with respect to β. These upperbounds imply that Assumption 2 is satisfied.

Note that the proofs above hold true for stationary and nonstationary y_t.

A.3. *Test and Asymptotic Distribution for Hamilton's Model*

We give the detailed calculation of the test statistic and its asymptotic distribution in Hamilton's model given in Example 3.4.

Under $H_0 : \mu_1 = 0$, the log-likelihood function is simply

$$\begin{aligned}\log L_t &= l_t \\ &= -\log\sqrt{2\pi} - \frac{1}{2}\log(\sigma^2) \\ &\quad - \frac{\left(y_t - \mu\left(1 - \sum_{i=1}^r \phi_i\right) - \sum_{i=1}^r \phi_i y_{t-i}\right)^2}{2\sigma^2}.\end{aligned}$$

The associated first and second derivatives of the log-likelihood function are as follows:

$$
\begin{aligned}
\frac{\partial l_t}{\partial \mu} &= \frac{e_t}{\sigma^2}\left(1-\sum_{i=1}^{r}\phi_i\right),\\
\frac{\partial l_t}{\partial \phi_i} &= \frac{e_t}{\sigma^2}(y_{t-i}-\mu),\\
\frac{\partial l_t}{\partial \sigma^2} &= -\frac{1}{2\sigma^2}+\frac{e_t^2}{2\sigma^4},\\
\frac{\partial^2 l_t}{\partial \mu^2} &= -\frac{1}{\sigma^2}\left(1-\sum_{i=1}^{r}\phi_i\right)^2,\\
\frac{\partial^2 l_t}{\partial \mu\,\partial \phi_i} &= -\frac{e_t}{\sigma^2}-\frac{y_{t-i}-\mu}{\sigma^2}\left(1-\sum_{i=1}^{r}\phi_i\right),\\
\frac{\partial_t^2 l_t}{\partial \mu\,\partial \sigma^2} &= -\frac{e_t}{\sigma^4}\left(1-\sum_{i=1}^{r}\phi_i\right),\\
\frac{\partial^2 l_t}{\partial \phi_i\,\partial \phi_j} &= -\frac{1}{\sigma^2}(y_{t-i}-\mu)(y_{t-j}-\mu),\\
\frac{\partial^2 l_t}{\partial \phi_i\,\partial \sigma^2} &= -\frac{e_t}{\sigma^4}(y_{t-i}-\mu),\\
\frac{\partial^2 l_t}{\partial \sigma^4} &= \frac{1}{2\sigma^4}-\frac{e_t^2}{\sigma^6},
\end{aligned}
$$

where $e_t = y_t - \mu(1-\sum_{i=1}^{r}\phi_i) - \sum_{i=1}^{r}\phi_i y_{t-i}$.

To test for $\mu_1 = 0$, the implied η_t is $(\mu_1 S_t\ \ \mathbf{0}\ \ 0)'$ following our notation. As a consequence,

$$
\operatorname{trace}\left(\left(l_t^{(2)}+l_t^{(1)}l_t^{(1)\prime}\right)E\left(\eta_t\eta_t'\right)\right) = c^2\frac{(e_t^2-\sigma^2)}{\sigma^4}\left(1-\sum_{i=1}^{r}\phi_i\right)^2,
$$

$$
\sum_{s<t}\operatorname{trace}\left(l_t^{(1)}l_s^{(1)\prime}E\left(\eta_t\eta_s'\right)\right) = \frac{c^2\left(1-\sum_{i=1}^{r}\phi_i\right)^2}{\sigma^4}\sum_{s<t}\rho^{t-s}e_t e_s,
$$

and

$$\mu_{2,t}^*(\beta,\theta_0)=\frac{\left(1-\sum_{i=1}^{r}\phi_i\right)^2}{2\sigma^4}\left(\left(e_t^2-\sigma^2\right)+2\sum_{s<t}\rho^{t-s}e_te_s\right).$$

To implement the test, we first estimate the MLE of parameters $\theta=(\mu,\phi_1,\dots,\phi_r,\sigma^2)'$ under H_0, denoted as $\hat{\theta}$. Then we regress $\mu_{2,t}(\beta,\hat{\theta})$ on $l_t^{(1)}(\hat{\theta})$ to obtain the residuals. Note that $\sum_t(\hat{e}_t^2-\hat{\sigma}^2)=0$, so that

$$\sum_t \mu_{2t}^*(\beta,\hat{\theta})=\frac{\left(1-\sum_{i=1}^{4}\hat{\phi}_i\right)^2}{2\hat{\sigma}^4}\sum_{s<t}\rho^{t-s}\hat{e}_t\hat{e}_s.$$

The asymptotic distribution can be computed using Theorem 4.3, namely, the asymptotic distribution of $\sum_t \mu_{2t}^*(\beta,\hat{\theta})/\sqrt{T}$ can be deducted from that of $\sum_t(\mu_{2t}^*(\beta,\theta_0)-d^{*\prime}l_t^{(1)}(\theta_0))/\sqrt{T}$, where $d^*=I(\theta_0)^{-1}\operatorname{cov}(\mu_{2t}^*(\beta,\theta_0),l_t^{(1)}(\theta_0))$. We derive this asymptotic distribution in two special cases where $r=0$ and $r=1$, respectively.

No AR Term

In this case, the formulas given above hold with $u_t=e_t$ and $\phi_1=\phi_2=\cdots=\phi_r=0$:

$$\begin{aligned}d^*&=I^{-1}\operatorname{cov}\left(\mu_{2,t}^*(\theta_0),l_t^{(1)}(\theta_0)\right)\\&=\frac{1}{2\sigma^4}\begin{bmatrix}\sigma^2&0\\0&2\sigma^4\end{bmatrix}\cdot\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}.\end{aligned}$$

Therefore,

$$\Gamma_T^*\overset{d}{=}\frac{1}{\sqrt{T}\sigma^4}\sum_t\sum_{s<t}\rho^{t-s}u_tu_s.$$

It follows from Andrews and Ploberger (1996) that, under H_0, Γ_T^* converges weakly to a linear combination of Gaussian processes:

$$\Gamma_T^*\Rightarrow\sum_{i=1}^{\infty}\rho^iZ_i, \tag{A.1}$$

where Z_i are i.i.d. standard Gaussian variables. Moreover, $\frac{1}{T}\sum \hat{\varepsilon}_t^{*2}$ converges to $\mathrm{Var}(\sum_{i=1}^{\infty} \rho^i Z_i) = \rho^2/(1-\rho^2)$. Hence,

$$\frac{\Gamma_T^*}{\sqrt{\frac{1}{T}\sum \hat{\varepsilon}_t^{*2}}} \Rightarrow \frac{\sqrt{1-\rho^2}\sum_{i=1}^{\infty}\rho^i Z_i}{|\rho|} = \begin{cases} \sqrt{1-\rho^2}\sum_{i=0}^{\infty}\rho^i Z_i, & \text{if } \rho > 0, \\ -\sqrt{1-\rho^2}\sum_{i=0}^{\infty}\rho^i Z_i, & \text{if } \rho < 0. \end{cases}$$

Note that, for $\rho = 0$, TS converges to 0 and hence supTS converges to 0 and expTS converges to 1. Let K denote the process $\mathrm{sign}(\rho)\sqrt{1-\rho^2}\sum_{i=0}^{\infty}\rho^i Z_i$, where $\mathrm{sign}(\rho) = 1$ if $\rho > 0$, $= 0$ if $\rho = 0$, and $= -1$ if $\rho < 0$. Then,

$$\text{(A.2)} \qquad \mathrm{supTS} \Rightarrow \sup_{\{\rho:\underline{\rho}<\rho<\bar{\rho}\}} \frac{1}{2}\big(\max(0, K)\big)^2,$$

and

$$\text{(A.3)} \qquad \mathrm{expTS} = \operatorname*{avg}_{\underline{\rho}\le\rho\le\overline{\rho}} \Psi(\rho) \quad \text{with}$$

$$\text{(A.4)} \qquad \Psi(\rho) \Rightarrow \begin{cases} \sqrt{2\pi}\exp\left[\frac{1}{2}(K-1)^2\right]\Phi(K-1), & \text{for } \rho \neq 0, \\ 1, & \text{for } \rho = 0. \end{cases}$$

From the continuous mapping theorem, we obtain the asymptotic distribution of expTS under H_0.

Case of an AR(1)

Now $r = 1$. Let us compute d^*:

$$E\left(\mu_{2t}^*\frac{\partial l_t}{\partial \mu}\right) = 0,$$

$$E\left(\mu_{2t}^*\frac{\partial l_t}{\partial \phi}\right) = \frac{(1-\phi)^2}{\sigma^6}\sum_{s<t}\rho^{t-s}E\big(e_t^2 u_{t-1} e_s\big).$$

Using the fact that $u_t = \phi^t u_0 + \phi^{t-1}e_1 + \cdots + e_t$, we have $E(u_{t-1}e_s) = \phi^{t-1-s}\sigma^2$ and hence

$$E\left(\mu_{2t}^*\frac{\partial l_t}{\partial \phi}\right) = \frac{(1-\phi)^2}{\sigma^2}\sum_{s<t}\rho^{t-s}\phi^{t-1-s}$$

$$= \frac{(1-\phi)^2}{\sigma^2\phi}\sum_{s<t}(\rho\phi)^{t-s}$$

$$= \frac{\rho(1-\phi)^2}{\sigma^2(1-\rho\phi)},$$

$$E\left(\mu_{2t}^* \frac{\partial l_t}{\partial \sigma^2}\right) = \frac{(1-\phi)^2}{4\sigma^8} E\big((e_t^2-\sigma^2)^2\big)$$

$$= \frac{(1-\phi)^2}{2\sigma^4}.$$

The information matrix is given by

$$I = -E\left(\frac{\partial^2 \log L}{\partial\theta\,\partial\theta'}\right) = \begin{bmatrix} (1-\phi)^2/\sigma^2 & 0 & 0 \\ 0 & 1/\big(1-\phi^2\big) & 0 \\ 0 & 0 & 1/\big(2\sigma^4\big) \end{bmatrix},$$

so that

$$d^* = \left(0, \frac{\rho(1-\phi^2)(1-\phi)^2}{\sigma^2(1-\rho\phi)}, (1-\phi)^2\right)',$$

$$\mu_{2t}^* - d^{*\prime} l_t^{(1)} = \frac{(1-\phi)^2}{\sigma^4} \sum_{s<t} \rho^{t-s} e_t e_s$$

$$- \frac{\rho(1-\phi^2)(1-\phi)^2}{\sigma^4(1-\rho\phi)} \sum_{s<t} \phi^{t-s-1} e_t e_s$$

$$= \frac{(1-\phi)^2}{\sigma^4}\left[\rho \sum_{s<t} \rho^{t-s-1} e_t e_s - \frac{\rho(1-\phi^2)}{(1-\rho\phi)} \sum_{s<t} \phi^{t-s-1} e_t e_s\right],$$

and from Andrews and Ploberger (1996), we have, under H_0,

$$\frac{1}{\sqrt{T}} \sum_t \big(\mu_{2t}^* - d^{*\prime} l_t^{(1)}\big) \Rightarrow \frac{(1-\phi)^2\rho}{\sigma^2}\left[\sum_{i=0}^{\infty} \rho^i Z_i - \frac{(1-\phi^2)}{(1-\rho\phi)} \sum_{i=0}^{\infty} \phi^i Z_i\right],$$

where Z_i are i.i.d. standard normal random variables. Moreover, under H_0:

$$\frac{\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*}{T} \xrightarrow{P} \frac{(1-\phi)^4\rho^2}{\sigma^4}\left[\frac{1}{1-\rho^2} - \frac{(1-\phi^2)}{(1-\rho\phi)^2}\right]$$

$$= \frac{(1-\phi)^4\rho^2}{\sigma^4} \frac{(\rho-\phi)^2}{(1-\rho^2)(1-\rho\phi)^2}.$$

We see that the variance is zero for $\phi = \rho$ and for $\rho = 0$. So for $\rho \neq 0$ and $\phi \neq \rho$, we have

$$\frac{\Gamma_T^*}{\sqrt{\frac{1}{T}\sum \hat{\varepsilon}_t^{*2}}} \Rightarrow \frac{\sqrt{1-\rho^2}|1-\rho\phi|}{|\rho-\phi|}\left[\sum_{i=0}^{\infty}\rho^i Z_i - \frac{(1-\phi^2)}{(1-\rho\phi)}\sum_{i=0}^{\infty}\phi^i Z_i\right].$$

A.4. *Asymptotic Critical Values for Model (6.1)*

Garcia's model, given in (6.1), corresponds to Hamilton's model with $r = 0$. The asymptotic distributions for supTS and expTS are given in (A.2) and (A.3). Now, we use 100,000 replications to tabulate the asymptotic critical values. We approximate the infinite sum in (A.1) and (A.3) by the finite sum truncated at $TR = 500$. A grid search of ρ is over the interval $[-0.7, 0.7]$ and $[-0.98, 0.98]$ with increment 0.01.[1] We also report the corresponding empirical critical values calculated for $T = 500$ with 10,000 iterations. See Table A-I.

The empirical critical values are very close to the asymptotic ones for both supTS and expTS tests, especially when $\rho \in [-0.7, 0.7]$. Note that the asymptotic critical values of supTS are much lower than those provided by Garcia and are also smaller than the cut-off points given by a $\chi^2(1)$, the distribution obtained in the standard case.

A.5. *Estimation of Hamilton's Model*

We estimate Hamilton's model given in Example 3.4 with $r = 4$ using Hamilton's original data set on U.S. GNP growth from 1952Q2 to 1984Q4, but also an extended series from 1952Q2 to 2010Q4.[2] The estimation under H_0 of linearity is as in Table A-II.

It is shown in Section D.4 below that the tests for $H_0 : \mu_1 = 0$ may lack power if ρ takes some specific values. To assess the power of our tests, we need to check the polynomial

$$\rho^4 - 0.310\rho^3 - 0.127\rho^2 + 0.121\rho + 0.089 = 0.$$

The roots to this polynomial are $0.524 \pm 0.415i$ and $-0.369 \pm 0.251i$. This implies that we cannot write $\mu_{2,t}$ as a linear combination of the first-order derivatives of the log-likelihood; consequently, our test has power. Similar results are obtained for the extended series, where all the roots are complex, too.

Using the extended series, we estimate the model under the alternative of an AR(4) with switching mean and variance and obtain Table A-III.

[1]Since $\rho = p + q - 1$, the interval $[-0.7, 0.7]$ for ρ corresponds to $p, q \in [0.15, 0.85]$ and the interval $[-0.98, 0.98]$ for ρ corresponds to $p, q \in [0.01, 0.99]$.

[2]The real GNP series is downloaded from St. Louis Fed's FRED database with id GNPC96. The variable used for y is 100 times the change in the log of real GNP.

TABLE A-I

EMPIRICAL AND ASYMPTOTIC CRITICAL VALUES OF SUPTS AND EXPTS

Percentile/c.v.	supTS Asymptotic	supTS Empirical	expTS Asymptotic	expTS Empirical
		$\rho \in [-0.7, 0.7]$		
99%	3.96	3.99	4.17	4.24
95%	2.45	2.51	1.82	1.87
90%	1.82	1.81	1.35	1.33
80%	1.21	1.19	1.04	1.02
70%	0.86	0.85	0.90	0.89
50%	0.45	0.43	0.76	0.75
10%	0.04	0.04	0.59	0.58
5%	0.01	0.01	0.55	0.55
1%	0.00	0.00	0.50	0.50
		$\rho \in [-0.98, 0.98]$		
99%	4.52	4.19	3.83	3.81
95%	2.99	2.78	1.82	1.76
90%	2.32	2.09	1.38	1.31
80%	1.65	1.44	1.07	1.02
70%	1.25	1.07	0.93	0.89
50%	0.74	0.60	0.78	0.74
10%	0.11	0.07	0.57	0.56
5%	0.05	0.03	0.54	0.52
1%	0.00	0.00	0.48	0.47

TABLE A-II

ML ESTIMATION OF GAUSSIAN AR(4) MODEL, U.S. REAL GNP

Parameters	1952Q2 to 1984Q4 Estimates	1952Q2 to 1984Q4 Standard Error	1952Q2 to 2010Q4 Estimates	1952Q2 to 2010Q4 Standard Error
μ	0.720	0.112	0.763	0.083
ϕ_1	0.310	0.085	0.335	0.076
ϕ_2	0.127	0.095	0.124	0.082
ϕ_3	−0.121	0.087	−0.083	0.074
ϕ_4	−0.089	0.090	−0.074	0.074
σ	0.983	0.061	0.883	0.056
T	131		235	

TABLE A-III

ML ESTIMATION OF TWO-STATE MARKOV SWITCHING AR(4) MODEL, U.S. REAL GNP

	1952Q2 to 2010Q4	
Parameters	Estimates	S.E.
μ	0.713	0.172
μ_1	0.099	0.232
ϕ_1	0.305	0.083
ϕ_2	0.210	0.083
ϕ_3	−0.120	0.077
ϕ_4	−0.053	0.064
σ	1.105	0.090
σ_1	0.421	0.056
p	0.956	0.031
q	0.967	0.031
L	279.142	

APPENDIX B: NOTATIONS

B.1. *Multilinear Forms*

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 would 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 (or "form," for short) of order p from V into F is a mapping M from $V \times \cdots \times V$ (where we take the product p times) to F, which is linear in each of the arguments. So

$$\lambda M\big(x^{(1)}, x^{(2)}, \ldots, x_1^{(i)}, \ldots, x^{(p)}\big) + \mu M\big(x^{(1)}, x^{(2)}, \ldots, x_2^{(i)}, \ldots, x^{(p)}\big) \tag{B.1}$$

$$= M\big(x^{(1)}, x^{(2)}, \ldots, \lambda x_1^{(i)} + \mu x_2^{(i)}, \ldots, x^{(p)}\big). \tag{B.2}$$

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

$$\lim_{r\to 0} \sup_{\|h\|=r} \big\| f(x+h) - f(x) - Df(x)(h) \big\| / r \to 0. \tag{B.3}$$

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. 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

$$x \to Df(x),$$

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 $D^2 f(x)$. 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 $h \in V$, $D^2 f(x)(h)$ is an element of $L(V, F)$, so for $k \in V$, $D^2 f(x)(h)(k)$ is an element of F. Moreover, 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.

When f has enough "derivatives," we can iterate this process and define the nth derivative $D^n f$ as derivative of $D^{n-1} f$,

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

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

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

with

$$R_n = \frac{1}{n!} \int_0^1 (1-t)^n D^{n+1} f(x+th)(h, \dots, h)\, dt, \tag{B.4}$$

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

Furthermore, provided that f is n times continuously differentiable, $D^n f$ is symmetric, that is,

$$D^n f(x)(h_1, \dots, h_n) = D^n f(x)(h_{\pi(1)}, \dots, h_{\pi(n)}) \tag{B.5}$$

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 nth derivative of g. Then

$$g^{(n)}(0) = D^n f(x)(h, \dots, h). \tag{B.6}$$

It is now an elementary, but tedious, exercise to show that, due to the symmetry (B.5), the multilinear form $D^n f(x)$ is uniquely defined by its values $D^n f(x)(h, \dots, h)$. (As an example, notice that, for a scalar bilinear form B, we have

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

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 $D(f \circ g) = Df(Dg)$. 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 stay with multilinear forms, and only define the operations and concepts we need. The experts will see that they are special cases of tensor algebra. Our 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).
- We are analyzing derivatives, so most of our multilinear forms are symmetric.

Assume that our reference, finite-dimensional vector space V is k-dimensional and that $b_1, \dots, b_k$ 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. For example, 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

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

We will now mainly work with *scalar* multilinear forms (i.e., the values of the form are real numbers). Hence we will assume—except when explicitly stated otherwise—that multilinear forms are scalar. Let now M be such a multilinear form. Then, using linearity, we have

$$M\big(x^{(1)}, x^{(2)}, \dots, x^{(p)}\big) = \sum M(b_{i_1}, \dots, b_{i_p}) x_{i_1}^{(1)} x_{i_2}^{(2)} \cdots x_{i_p}^{(p)}, \tag{B.7}$$

where the sum symbol corresponds to p sums extending over all values of $i_1, \dots, i_p$ between 1 and k. There is a one-to-one correspondence between the k^p numbers $M(b_{i_1}, \dots, b_{i_p})$ 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, having fixed the coordinate system, we can *identify* the multilinear form M with its coordinates $M(b_{i_1}, \dots, b_{i_p})$. Multilinear forms (with the usual operations) of order p form a finite-dimensional vector space. The only difference to a "usual" vector space is the enumeration of the coordinates. We do not index them by the numbers of $1, \dots, K$, but our index set consists of the p-tuples $(1, \dots, 1), (2, 1, \dots), \dots, (k, k, \dots, k)$. Note that bilinear forms (forms of order 2) are $k \times k$-matrices.

This way, we can work with multilinear forms and related mathematical objects without having to discuss tensor algebra:

1. Multilinear forms form a vector space, and the mapping attaching each multilinear form to its coordinates is an isomorphism. Hence we do not need to distinguish between multilinear forms and k^p numbers indexed by a multi-index $(i_1, \dots, i_p)$.

2. Let us call a multilinear form C defined by coordinates $(c_{i_1,\dots,i_p})$ *symmetrical* if and only if, for all $(i_1, \dots, i_p)$ and all permutations π of numbers between 1 and k,

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

This property is equivalent to our definition above, (B.5). For a form C defined by coordinates $(c_{i_1,\dots,i_p})$, define its symmetrization $C^{(S)}$ by

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

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

$$C(h, \dots, h) = C^{(S)}(h, \dots, h), \tag{B.8}$$

and, for any form C, $C^{(S)}$ is the only symmetrical form with the property (B.8).

3. Another special case of multilinear forms are our derivatives of scalar functions defined on open subsets of our space V. The coordinates $D^n f$ can be calculated in the following way. Define the function g by

$$g(x_1, \dots, x_p) = f\Big(\sum x_i b_i\Big), \tag{B.9}$$

where the b_i are our fixed basis vectors. Then the corresponding coordinates of the derivative are given by $(\frac{\partial^n g}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_n}})_{(i_1,\dots,i_n)}$.

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

$$g_h(t) = f(x + th).$$

Then, it follows from (B.6) that $D^n f(h, h, \dots, h) = g_h^{(n)}(0)$, where $g_h^{(n)}$ is the usual nth derivative. Now suppose we can find a form C so that, for all h,

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

Then, due to (B.8) and the symmetry of the derivative, $D^n f = C^{(S)}$.

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

$$a_{(i_1,\dots,i_p)} b_{(i_{p+1},\dots,i_{p+q})}. \tag{B.10}$$

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 forms of order 2 is a form of order 4. It is interesting to consider the properties of the corresponding multilinear forms:

$$(A \otimes B)(h_1, \dots, h_{p+q}) = A(h_1, \dots, h_p) B(h_{p+1}, \dots, h_{p+q}). \tag{B.11}$$

The tensor product of symmetric forms, however, may not be symmetric.

6. We can define the scalar product $\langle \cdot, \cdot \rangle$ in the usual way. Let us assume that T represents a form with coordinates $(t_{i_1,\dots,i_p})$, and C is a form with coordinates $(c_{i_1,\dots,i_p})$; we have

$$\langle T, C \rangle = \sum t_{i_1,\dots,i_p} c_{i_1,\dots,i_p}. \tag{B.12}$$

7. This scalar product is useful in computing the expectation of multilinear forms with random arguments. First of all, let us observe that each vector $h \in V$ has exactly k coordinates. Since (B.7) defines, for each set of coordinates, a form, we can identify h with a 1-form (i.e., a linear form with one argument).

We will use the same symbol h for this form. Now let $h_1, \dots, h_p \in V$. Then we can use (B.10) to define $h_1 \otimes \cdots \otimes h_p$. Now suppose we want to compute the value of the multilinear form $T(h_1, \dots, h_p)$. Then we can see from (B.7), (B.12) that $T(h_1, \dots, h_p)$ equals $\langle T, h_1 \otimes \cdots \otimes h_p \rangle$. Let $H_1, \dots, H_p$ be random variables with values in our reference space V, and T be a multilinear form, which is fixed or exogenous. Suppose we want to compute the expectation of

$$T(H_1, \dots, H_p).$$

Since $T(H_1, \dots, H_p) = \langle T, H_1 \otimes \cdots \otimes H_p \rangle$, and since T is independent of the H_i, we can have

$$E\big(T(H_1 \otimes \cdots \otimes H_p)\big) = \big\langle T, E(H_1 \otimes \cdots \otimes H_p)\big\rangle, \tag{B.13}$$

provided the expectations exist. (A sufficient condition is, e.g., $E\|H_1\| \cdots \|H_p\| < \infty$: $H_1 \otimes \cdots \otimes H_p$ is a multilinear form, and, as already mentioned above, the forms of order p form a vector space. Hence we should not have any conceptual difficulties with expectations.) Moreover, (B.13) is valid for conditional expectations, too. In the sequel, we will use these types of identities rather freely.

8. Most of our proof will involve the expectation of multilinear forms representing derivatives. The notation using the bracket $\langle \cdot, \cdot \rangle$ would be rather clumsy. So we propose to use a more suggestive notation. Instead of $\langle T, C \rangle$, we will use $T(C)$, that is, we use the form C as an argument. With this notation, we can write (B.13) as

$$E\big(T(H_1 \otimes \cdots \otimes H_p)\big) = T\big(E(H_1 \otimes \cdots \otimes H_p)\big).$$

Furthermore, when evaluating these kinds of expressions, we will use the usual linearity properties of scalar products without further notice.

9. If A is symmetrical, then, for every T,

$$T(A) = T^{(S)}(A). \tag{B.14}$$

In particular, if we have an arbitrary random vector H (with sufficiently many moments), then $E(H \otimes \cdots \otimes H)$ is symmetrical, hence (B.14) implies that, for all forms T,

$$T\big(E(H \otimes \cdots \otimes H)\big) = T^{(S)}\big(E(H \otimes \cdots \otimes H)\big).$$

10. As we already stated, the multilinear forms 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 about which norm we use. Of particular interest, however, is the norm

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

where the $t_{i_1,\dots,i_p}$ are the coordinates of T. The Cauchy–Schwarz inequality and (B.12) imply that, for all T, C,

$$\left|T(C)\right| \le \|T\|\|C\|.$$

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

B.2. *Other Notations*

The sample is split into blocks in the following way:

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

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

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

In the proofs, we choose B_L so that some terms become negligible.

DEFINITION B.1: Define $\mathcal{H}_{i,T}$ as the σ-algebra generated by $(\vartheta_{T_i}, \vartheta_{T_i-1}, \dots, \vartheta_1, y_T, \dots, y_1)$, where ϑ_t was introduced in Assumption 3.

Then $\mathcal{H}_{0,T}$ is the σ-algebra generated by the data $(y_T, \dots, y_1)$ only.

Our analysis is based on the derivatives of the logarithm of the likelihood function. Recall that the conditional densities are denoted by $f_t = f_t(\theta_T)$, and the conditional log-likelihood functions by l_t. We also defined $D^k l_t = l_t^{(k)}$.

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

$$\ell_t(u) = \log f_t(\theta_T + uh).$$

Let $f = f_t(\theta_T)$ and $f', f^{(2)}, \dots$ denote the derivatives of $f_t(\theta_T + uh)$ with respect to u. When differentiating ℓ_t, one obtains:

$$\begin{aligned}
&\text{1st derivative:} \quad \ell_t^{(1)} = \frac{f'}{f},\\
&\text{2nd derivative:} \quad \ell_t^{(2)} = \frac{f^{(2)}}{f} - \frac{f'}{f^2} f',\\
&\text{3rd derivative:} \quad \ell_t^{(3)} = \frac{f^{(3)}}{f} - \frac{f^{(2)}}{f^2} f' - \frac{2f' f^{(2)}}{f^2} + 2\frac{f'^2}{f^3} f',\\
&\text{4th derivative:} \quad \ell_t^{(4)} = \frac{f^{(4)}}{f} - \frac{f^{(3)}}{f^2} f' - \frac{3f^{(3)} f' + 3f^{(2)} f^{(2)}}{f^2} + 6\frac{f^{(2)} f'}{f^3} f'\\
&\qquad\qquad + 6\frac{f'^2}{f^3} f^{(2)} - 6\frac{f'^3}{f^4} f'.
\end{aligned}$$

According to the formalism outlined previously, we can conclude that $\ell_t^{(k)} = l_t^{(k)}(h, \dots, h)$ and that $f^{(k)} = D^k f(h, \dots, h)$. Taking into account our characterization of the tensor product (B.11), and the techniques described above, we can conclude that

$$\begin{aligned}
&l_t^{(1)} = (1/f_t) D f_t,\\
&\frac{1}{f_t} D^2 f_t = l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)},\\
&\frac{1}{f_t} D^3 f_t = \left(l_t^{(3)} + 3 l_t^{(2)} \otimes l_t^{(1)} + l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)}\right)^{(S)},\\
&\frac{1}{f_t} D^4 f_t = \left(l_t^{(4)} + 6 l_t^{(2)} \otimes l_t^{(1)} \otimes l_t^{(1)} + 4 l_t^{(3)} \otimes l_t^{(1)}\right.\\
&\qquad\qquad \left. + 3 l_t^{(2)} \otimes l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)}\right)^{(S)}.
\end{aligned} \tag{B.15}$$

There is no need to symmetrize (B.15) since the form on the right hand side is symmetrical. Moreover, for $k \leq 4$, we have, for arbitrary h, $E(\frac{1}{f_t} D^k f_t(h, \dots, h) | \mathcal{H}_{0,t-1}) = \int \frac{1}{f_t} D^k f_t(h, \dots, h) f_t \, d\mu(y_t) = \int D^k f_t(h, \dots, h) \, d\mu(y_t)$, where μ is the dominating measure defined in Section 2. Since we assumed f_t to be at least five times differentiable (and the fifth derivative to be uniformly integrable), it follows from Bartle (1966, Corollary 5.9) that we can interchange integral and differentiation, and conclude that $\int D^k f_t(h, \dots, h) \, d\mu(y_t) = D^k(\int f_t \, d\mu(y_t))(h, \dots, h) = 0$, since all the f_t, as conditional densities, integrate to 1. It follows that $E(\frac{1}{f_t} D^k f_t(h, \dots, h) | \mathcal{H}_{0,t-1}) = 0$. This property is referred to as the kth Bartlett identity. Note that as a consequence, $\frac{1}{f_t} D^k f_t(h, \dots, h)$ is a martingale difference sequence with respect to $\mathcal{H}_{0,t}$.

APPENDIX C: PROOFS

C.1. *Proofs of Theorems 3.1 and 3.2*

The proof of Theorem 3.1 uses the following lemma.

LEMMA C.1: *We have*

$$\frac{1}{T}\sum_t \frac{\partial \mu_{2,t}(\beta,\theta_0)}{\partial\theta} = -\frac{1}{T}\sum_t \mu_{2,t}(\beta,\theta_0)\frac{\partial l_t(\theta_0)}{\partial\theta} + o_p(1),$$

uniformly in β.

PROOF: The proof consists in some manipulations of the derivatives of $\mu_{2,t}$ with respect to θ. To make our formulas more readable, we omit all the arguments like $E(\eta_t \otimes \eta_s)$ or similar moments of η_t. They only depend on β, so all differentiation with respect to θ will leave them unchanged. Let us first rewrite $\frac{\partial\mu_{2,t}}{\partial\theta_k}$. We have

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

and

$$\begin{aligned}\frac{\partial}{\partial\theta_k}\left(l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)}\right) &= \frac{\partial}{\partial\theta_k}\left(\frac{\partial^2 l_t}{\partial\theta_i\,\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial l_t}{\partial\theta_j}\right)\\ &= \frac{\partial^3 l_t}{\partial\theta_k\,\partial\theta_i\,\partial\theta_j} + \frac{\partial^2 l_t}{\partial\theta_k\,\partial\theta_i}\frac{\partial l_t}{\partial\theta_j} + \frac{\partial l_t}{\partial\theta_i}\frac{\partial^2 l_t}{\partial\theta_k\,\partial\theta_j}.\end{aligned}$$

From the third Bartlett identity,

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

is a martingale difference sequence (m.d.s.) and therefore $\frac{1}{T}\sum_{t=1}^{T} m_{3,t} = o_p(1)$. Of course, $m_{3,t}$ is still a function of β. But this parameter only appears in the arguments of the linear forms on the right hand side, which are moments of η_t. Since we assumed the η_t to be uniformly bounded and exponentially mixing, we can conclude that this relationship holds uniformly in β. We can use the same argument in all of the convergence results in this proof:

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

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

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

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

where $\widehat{\operatorname{cov}}$ denotes the empirical covariance. Moreover, we have $\widehat{\operatorname{cov}}(\mu_{2,t},\frac{\partial l_t}{\partial\theta_k})\to\operatorname{cov}(\mu_{2,t},\frac{\partial l_t}{\partial\theta_k})$. *Q.E.D.*

PROOF OF THEOREM 3.1: First let us observe that because $\hat{\theta}$ is the MLE, $\sum l_t^{(1)}(\hat{\theta}) = 0$. Hence Γ_T can be rewritten as $\frac{1}{\sqrt{T}} \sum(\mu_{2,t}(\beta, \hat{\theta}) - d' l_t^{(1)}(\hat{\theta}))$. Denote

$$\nu_T(\beta, \theta) = \frac{1}{\sqrt{T}} \sum_t \bigl(\mu_{2,t}(\beta, \theta) - d' l_t^{(1)}(\theta)\bigr). \tag{C.1}$$

To show point (i) of Theorem 3.1, we need to establish that (a) $\nu_T(\beta, \hat{\theta}) - \nu_T(\beta, \theta_0)$ converges to zero in probability uniformly in β, and (b) $\nu_T(\beta, \theta_0)$ converges to $N(\beta)$ as a process indexed by β.

(a) A second-order Taylor expansion of $\nu_T(\beta, \hat{\theta})$ around θ_0 gives

$$\begin{aligned} \nu_T(\beta, \hat{\theta}) &= \nu_T(\beta, \theta_0) + \frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} \nu_T(\beta, \theta_0) \sqrt{T}(\widehat{\theta} - \theta_0) \\ &\quad + \frac{1}{2} (\widehat{\theta} - \theta_0)' \frac{1}{\sqrt{T}} \frac{\partial^2}{\partial \theta \, \partial \theta'} \nu_T(\beta, \bar{\theta}) \sqrt{T}(\widehat{\theta} - \theta_0), \end{aligned}$$

where $\bar{\theta} = \lambda \theta_0 + (1 - \lambda)\hat{\theta}$ for some $0 < \lambda < 1$. Assumption 2 guarantees the uniform convergence (in θ) of $\frac{1}{\sqrt{T}} \frac{\partial^2}{\partial \theta \partial \theta'} \nu_T(\beta, \theta)$ to $E^{\theta_0}(\frac{1}{\sqrt{T}} \frac{\partial^2}{\partial \theta \partial \theta'} \nu_T(\beta, \theta))$, which is a constant. As $\sqrt{T}(\widehat{\theta} - \theta_0) = O_p(1)$, it is sufficient to show that

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

in probability uniformly in β. To establish (C.2), we have to show that

$$\frac{1}{T} \sum_t \frac{\partial \mu_{2,t}(\beta, \theta_0)}{\partial \theta} - \frac{1}{T} \sum_{t=1}^{T} d' l_t^{(2)}(\theta_0) \xrightarrow{P} 0,$$

uniformly in β. The average of the second derivatives converges to the negative Information matrix,

$$\frac{1}{T} \sum_{t=1}^{T} l_t^{(2)}(\theta_0) \xrightarrow{P} -I(\theta_0),$$

and from Lemma C.1, it follows that

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

Then (C.2) follows from the definition of d in (3.1).

(b) By Pollard (1990), ν_T converges weakly to a process ν if (i) B is totally bounded, (ii) the finite-dimensional distributions of ν_T converge to those of ν, and (iii) $\{\nu_T(\theta_0,\cdot): T \geq 1\}$ is stochastic equicontinuous, that is, for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\overline{\lim}_{T\to\infty} P\Bigl(\sup_{\|\beta_1-\beta_2\|\leq\delta} \bigl|\nu_T(\theta_0,\beta_1) - \nu_T(\theta_0,\beta_2)\bigr| > \varepsilon\Bigr) < \varepsilon.$$

Condition (i) holds because B is a compact. Condition (ii) holds by the central limit theorem of martingale difference sequences. We now establish (iii). The process $\frac{1}{\sqrt{T}}\sum_t \mu_{2,t}(\theta,\beta)$ can be approximated by

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_t \tilde{\mu}_{2,t}(\theta,\beta) &= \frac{1}{2\sqrt{T}}\sum_{t=1}^{T} \mathrm{tr}\Biggl(\bigl((l_t^{(2)} + l_t^{(1)} l_t^{(1)\prime}) E^\beta(\eta_t\eta_t')\bigr) \\
&\quad + 2\sum_{s=-\infty}^{t-1} \bigl(l_t^{(1)} l_s^{(1)\prime} E^\beta(\eta_t\eta_s')\bigr)\Biggr) \\
&= \frac{1}{2}\mathrm{tr}\Biggl(E^\beta(\eta_t\eta_t')\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(l_t^{(2)} + l_t^{(1)} l_t^{(1)\prime}) \\
&\quad + 2\sum_{k=1}^{\infty} E^\beta(\eta_t\eta_{t-k}')\frac{1}{\sqrt{T}}\sum_t (l_t^{(1)} l_{t-k}^{(1)\prime})\Biggr) \\
&= \frac{1}{2}\mathrm{tr}\bigl(E^\beta(\eta_t\eta_t')m_0\bigr) + \mathrm{tr}\Biggl(\sum_{k=1}^{\infty} E^\beta(\eta_t\eta_{t-k}')m_k\Biggr),
\end{aligned}$$

where $m_0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(l_t^{(2)} + l_t^{(1)} l_t^{(1)\prime})$ and $m_k = \frac{1}{\sqrt{T}}\sum_t (l_t^{(1)} l_{t-k}^{(1)\prime})$ are both $O_p(1)$. Let us denote by s_T the term $\frac{1}{\sqrt{T}}\sum_t l_t^{(1)}(\theta_0)$. We can approximate $\nu_T(\theta_0,\cdot)$ by

$$\widetilde{\nu}_T(\beta) = \frac{1}{2}\mathrm{tr}\bigl(E^\beta(\eta_t\eta_t')m_0\bigr) + \mathrm{tr}\Biggl(\sum_{k=1}^{\infty} E^\beta(\eta_t\eta_{t-k}')m_k\Biggr) - d'(\beta)s_T.$$

We have

$$\begin{aligned}
E\sup_{\beta\in B}\bigl|\nu_T(\beta) - \widetilde{\nu}_T(\beta)\bigr| &\leq E\sup_{\beta\in B}\Biggl|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{s=-\infty}^{0}\mathrm{tr}\bigl(l_t^{(1)} l_s^{(1)\prime} E^\beta(\eta_t\eta_s')\bigr)\Biggr| \\
&\leq \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{s=-\infty}^{0} E\sup_{\beta\in B}\bigl|\mathrm{tr}\bigl(l_t^{(1)} l_s^{(1)\prime} E^\beta(\eta_t\eta_s')\bigr)\bigr|
\end{aligned}$$

$$\leq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{s=-\infty}^{0} E\big(\big|l_t^{(1)}\big|\big|l_s^{(1)}\big|'\big) \sup_{\beta \in B} \big|E^{\beta}\big(\eta_t \eta_s'\big)\big|$$

$$\leq E\big(l_t^{(1)} l_t^{(1)\prime}\big) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{s=-\infty}^{0} \sup_{\beta \in B} \big|E^{\beta}\big(\eta_t \eta_s'\big)\big|$$

$$\leq E\big(l_t^{(1)} l_t^{(1)\prime}\big) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{s=-\infty}^{0} \lambda^{t-s} \sup_{\beta \in B} C^{\beta},$$

with $0 < \lambda < 1$, where the first inequality follows from Cauchy–Schwarz and the second follows from Assumption 3 (geometric ergodicity of η_t). Moreover, since C^{β} is some moment of η_t, it is necessarily finite by Assumption 3. Therefore, $E \sup_{\beta \in B} |\nu_T(\beta) - \widetilde{\nu}_T(\beta)|$ converges to zero. Thus, it suffices to establish the stochastic equicontinuity of $\{\widetilde{\nu}_T(\cdot) : T \geq 1\}$. We have

$$\begin{aligned}
\big|\widetilde{\nu}_T(\beta_1) - \widetilde{\nu}_T(\beta_2)\big| \leq{}& \Bigg|\operatorname{tr}\bigg(\big(E^{\beta_1}\big(\eta_t \eta_t'\big) - E^{\beta_2}\big(\eta_t \eta_t'\big)\big) \frac{m_0}{2}\bigg) \\
&+ \operatorname{tr}\Bigg(\sum_{k=1}^{\infty} \big(E^{\beta_1}\big(\eta_t \eta_{t-k}'\big) - E^{\beta_2}\big(\eta_t \eta_{t-k}'\big)\big) m_k\Bigg) \\
&- \big(d(\beta_1) - d(\beta_2)\big)' s_T\Bigg| \\
\leq{}& p\bigg\|\bigg(\big(E^{\beta_1}\big(\eta_t \eta_t'\big) - E^{\beta_2}\big(\eta_t \eta_t'\big)\big) \frac{m_0}{2}\bigg)\bigg\| \\
&+ p\Bigg\|\sum_{k=1}^{\infty} \big(E^{\beta_1}\big(\eta_t \eta_{t-k}'\big) - E^{\beta_2}\big(\eta_t \eta_{t-k}'\big)\big) m_k\Bigg\| \\
&+ \big|\big(d(\beta_1) - d(\beta_2)\big)' s_T\big|.
\end{aligned}$$

The second term is split into two sums so that

$$\begin{aligned}
&\Bigg\|\sum_{k=1}^{\infty} \big(E^{\beta_1}\big(\eta_t \eta_{t-k}'\big) - E^{\beta_2}\big(\eta_t \eta_{t-k}'\big)\big) m_k\Bigg\| \\
&\quad \leq \Bigg\|\sum_{k=1}^{k_0-1} \big(E^{\beta_1}\big(\eta_t \eta_{t-k}'\big) - E^{\beta_2}\big(\eta_t \eta_{t-k}'\big)\big) m_k\Bigg\| \\
&\qquad + \Bigg\|\sum_{k=k_0}^{\infty} \big(E^{\beta_1}\big(\eta_t \eta_{t-k}'\big) - E^{\beta_2}\big(\eta_t \eta_{t-k}'\big)\big) m_k\Bigg\|.
\end{aligned}$$

Assumption 3 implies that the second term is of order λ^{k_0} and hence can be made as small as we want by choosing k_0 sufficiently large. Finally, the stochastic equicontinuity follows from the continuity of $E^{\beta}(\eta_t \eta'_{t-k})$, $k = 0, 1, \ldots$ and of $d(\beta)$ and the application of the Markov theorem.

Point (ii) of Theorem 3.1. Let

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

Assumption 2 guarantees the uniform convergence in θ of $\frac{1}{T}\sum_{t=1}^{T} l_t^{(1)}(\theta) l_t^{(1)\prime}(\theta)$ to $E(l_t^{(1)}(\theta) l_t^{(1)\prime}(\theta))$. Similarly, it can be shown that $\frac{1}{T}\sum_{t=1}^{T} \mu_{2,t}(\theta, \beta) l_t^{(1)}(\theta)$ converges uniformly in β and θ to $E(\mu_{2,t}(\theta, \beta) l_t^{(1)}(\theta))$. Then, by the consistency of the ML estimator, it follows that $\hat{d}(\beta)$ converges to $d(\beta)$ in probability uniformly in β. Denote $y_t = \mu_{2,t}(\hat{\theta})$, $x_t = l_t^{(1)}(\hat{\theta})$, $y = (y_1, \ldots, y_T)'$, and $X = (x'_1, \ldots, x'_T)'$. Using these notations, $\hat{d} = (X'X)^{-1}X'y$ and

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

where $M_X = I - X(X'X)^{-1}X'$ is idempotent. Consequently,

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

in probability uniformly in β.

Point (iii) of Theorem 3.1 is an immediate consequence of points (i) and (ii) and the continuous mapping theorem. This concludes the proof of Theorem 3.1. *Q.E.D.*

PROOF OF THEOREM 3.2: First of all, let us state some high-level sufficient conditions.

CONDITION A: For each θ_0 and $\varepsilon > 0$, we can find an $M(\varepsilon)$ so that there exist random variables $W_1, \ldots, W_{M(\varepsilon)}$ with the following properties:

A1. There exists a continuous function f so that

$$P_{\theta_0+h/\sqrt{T}}\big[\big|\mathrm{TS} - f(W_1, \dots, W_{M(\varepsilon)})\big| > \varepsilon\big] < \varepsilon,$$

uniformly on all bounded subsets of h.

A2. For $T \to \infty$, $(W_1, \dots, W_{M(\varepsilon)}, \log \frac{dP_{\theta_0+h/\sqrt{T}}}{dP_{\theta_0}})'$ converges in distribution under P_{θ_0} to a Gaussian distribution with mean 0 and covariance

$$\begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix},$$

where Σ is the asymptotic covariance of $(W_1, \dots, W_{M(\varepsilon)})'$ and Ω the covariance matrix of $\log \frac{dP_{\theta_0+h/\sqrt{T}}}{dP_{\theta_0}}$.

Condition A1 stipulates that the test statistic, TS, can be approximated by a sequence of random variables. Condition A2 stipulates that the joint distribution of these random variables and the log-likelihood ratio is asymptotically normal with correlation equal to zero.

THEOREM C.2: *Consider any test that satisfies Conditions* A1 *and* A2. *The critical values obtained using data drawn from the distribution $P_{\hat{\theta}}$ are asymptotically the same as those obtained from data drawn from the distribution P_{θ_0}.*

PROOF: We apply Le Cam's third lemma (van der Vaart (1998, p. 90)) on Condition A2. It implies that, under P_{θ_T}, $(W_1, \dots, W_{M(\varepsilon)})'$ converges in distribution to a Gaussian distribution with mean 0 and covariance Σ. So the distribution is the same under the null and the local alternative. If we draw the observations in $P_{\hat{\theta}}$, the corresponding critical values will be asymptotically the same as if the observations were drawn in P_{θ_0}. *Q.E.D.*

PROOF OF THEOREM 3.2—Continued: We have to verify both Conditions A1 and A2 for our test statistic. First of all, let us observe that we are establishing some kind of pivotal property of our test statistic. $\mathrm{TS}_T(\beta, \hat{\theta})$ is a function of the data alone, so its distribution is determined by the underlying distribution of the data. We established in the proof of Theorem 3.1 that, under P_{θ_0}, the processes $\mathrm{TS}_T(\cdot, \hat{\theta})$ converge for $T \to \infty$ in distribution. Hence their probability distributions remain uniformly tight. For every $\varepsilon > 0$, we can find a compact set $C(\varepsilon)$ of continuous functions so that the probabilities of $\mathrm{TS}_T(\cdot, \hat{\theta})$ being in $C(\varepsilon)$ are at least $1 - \varepsilon$. The Arzela–Ascoli theorem characterizes the elements of compact sets to be equicontinuous. Equicontinuity implies that we can approximate (in the sense that, for all $\varepsilon > 0$, the probability of the difference being larger than ε becoming arbitrarily small) the integrals $\int \exp(\mathrm{TS}_T(\beta, \hat{\theta}))\, d\nu(\beta, d)$ by finite sums $\sum \nu_i \exp(\mathrm{TS}_T(\beta_i, \hat{\theta}))$. An analogous result holds true for the supremum statistic: Equicontinuity implies that $\sup(\mathrm{TS}_T(\beta, \hat{\theta}))$ can be approximated by $\max(\mathrm{TS}_T(\beta_i, \hat{\theta}))$.

Hence it is sufficient to show that the joint distributions of the finite-dimensional vectors $(\mathrm{TS}_T(\beta_i, \hat{\theta}) : 1 \leq i \leq N)$ are asymptotically the same normal distribution for all P_θ with θ such that $\|\theta - \theta_0\| \leq M/\sqrt{T}$ for arbitrary M. Condition A1 shows that, asymptotically, the density between probabilities corresponding to parameters $\theta_0 + h/\sqrt{T}$ and $\theta_0 + k/\sqrt{T}$ is lognormal with mean $O(\|h - k\|)$ and variance $O(\|h - k\|^2)$. Hence compactness guarantees that, for every $\varepsilon > 0$, we can find finitely many parameter values, say $h_1, \dots, h_j$, independent of T, 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 $(\mathrm{TS}_T(\beta_i, \hat{\theta}) : 1 \leq i \leq N)$ are the same when the data are generated by $\theta_0 + h_i/\sqrt{T}$.

To show this, we can apply Theorem 3.1. Under P_{θ_0}, the $\mathrm{TS}_T(\beta_i, \hat{\theta})$ are normalized sums of martingale-differences (plus constants), and elementary calculations establish that

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

Hence the multivariate CLT implies that the joint distribution of $\mathrm{TS}_T(\beta_i, \hat{\theta})$ and the logarithm of the densities is a multivariate normal distribution. Moreover, as $\mathrm{TS}_T(\beta_i, \hat{\theta})$ involves a projection on the space orthogonal to $l_t^{(1)}$, it is asymptotically uncorrelated and hence independent from the logarithm of the densities. This proves Condition A2. The consistency of fractiles follows from the fact that TS^s are i.i.d. across s, conditional on the observations. *Q.E.D.*

C.2. *Proof of Theorem 4.1: Preliminary*

Uniform convergence: The statement of the theorem involves some uniform convergence in probability of a parameterized family of random variables. First assume the theorem would not be true. There would exist 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 an $\varepsilon > 0$ so that

$$P_{\theta_T}\Bigg(\Bigg[\Bigg|\ell_T^{\beta_T}(\theta_T) \Big/ \exp\Bigg(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t}(\theta_T, \beta_T) - \frac{1}{2} E\big(\mu_{2,t}(\theta_T, \beta_T)^2\big)\Bigg) - 1\Bigg| \geq \varepsilon\Bigg]\Bigg) \geq \varepsilon.$$

Since the (θ_T, β_T) are elements of a compact subset, there exists a convergent subsequence. Hence, to prove Theorem 4.1, it is sufficient to show that, for every $(\theta_T, \beta_T) \to (\theta_0, \beta_0)$,

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

or

$$\ell_T^{\beta_T}(\theta_T)\Big/\exp\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mu_{2,t}(\theta_T,\beta_T) - \frac{1}{2}E\big(\mu_{2,t}(\theta_T,\beta_T)^2\big)\right) \to 1,$$

in probability with respect to P_{θ_T}.

In the sequel, we will prove this relationship. To simplify our notation, however, we 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 drop the argument from expressions like $l_t(\theta_T)$, and simply use l_t. The proper argument should be evident from the context. This simplification of notation 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 $l_t^{(2)}$ denotes the second derivative of l_t at θ_T, which is a bilinear form, and $l_t^{(2)}(h, k)$ is the evaluation of this bilinear form with the arguments h and k.

Reference spaces: In our construction of the alternative, we did assume that the parameters ϑ_t are strictly exogenous. Assume our random variables $(y_1, y_2, \ldots, y_T, \vartheta_1, \ldots, \vartheta_T)$ are defined on the product space

$$\Pi = \Omega^T \times \Xi^T \tag{C.3}$$

with σ-algebras

$$\mathcal{H}_{0,T} \times \sigma(\vartheta_1, \ldots, \vartheta_T),$$

and $\mathcal{H}_{0,T}$ is defined in Definition B.1. Moreover, under the null hypothesis, the probability measures on $\mathcal{H}_{0,T}$ and $\sigma(\vartheta_1, \ldots, \vartheta_T)$ are independent by construction (see Assumption 1). Hence

$$P = P|_{\mathcal{H}_{0,T}} \times P|_{\sigma(\vartheta_1,\ldots,\vartheta_T)},$$

where $P|.$ denotes the restriction of the measure to the σ-algebra. Remark that this independence is not true under the alternative. However, this does not matter because we only analyze the properties of the likelihood under the null. Since all our computations only involve the observations and functions of ϑ_t, we can, without limitation of generality, assume that all random variables are defined on the space Π as in (C.3).

Steps of the Proof: Theorem 4.1 is proved in three steps.

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

$$\mathrm{TE}_T = \sum_{t=1}^{T}\biggl[\frac{1}{\sqrt[4]{T}} l_t^{(1)}(\eta_t) + \frac{1}{2\sqrt{T}} l_t^{(2)}(\eta_t, \eta_t) + \frac{1}{6\sqrt[4]{T^3}} l_t^{(3)}(\eta_t, \eta_t, \eta_t) + \frac{1}{24T} l_t^{(4)}(\eta_t, \eta_t, \eta_t, \eta_t)\biggr],$$

where $l_t^{(1)}, \dots, l_t^{(4)}$ are function of θ_T.

Denote

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

Step 1. Show that

$$\text{(C.4)} \qquad \frac{\ell_T^{\beta_T}(\theta_T)}{E[\exp(\mathrm{TE}_T) \mid \mathcal{H}_{0,T}]} \xrightarrow{P} 1.$$

Step 2. Show that there exists some $\mathcal{H}_{0,T}$-measurable Z_T such that

$$\text{(C.5)} \qquad \lim E\bigl(\exp(\mathrm{TE}_T - Z_T) \mid \mathcal{H}_{0,T}\bigr) = 1.$$

Step 3. Show that

$$\text{(C.6)} \qquad Z_T - \widetilde{\mathrm{TS}}_T(\beta_T, \theta_T) = o_p(1).$$

The following lemmas are used in the proof of Theorem 4.1.

LEMMA C.3: *Assume that, for any* $\varepsilon > 0$, *we can find* $1 - \varepsilon \le \frac{f_T}{f_T^*} \le 1 + \varepsilon$ *on some set* A_T^ε *so that* $\lim_{T\to\infty} P(A_T^\varepsilon) = 1$, *where* A_T^ε *is* $\mathcal{H}_{0,T}$-*measurable and independent of* β. *Then* $\frac{E(f_T|\mathcal{H}_{0,T})}{E(f_T^*|\mathcal{H}_{0,T})} \xrightarrow{P} 1$.

Note that a sufficient condition for Lemma C.3 is

$$\left|\frac{f_T}{f_T^*}\right| \le 1 + C_T,$$

where C_T is $\mathcal{H}_{0,T}$-measurable and independent of β and $C_T \xrightarrow{p} 0$.

PROOF OF LEMMA C.3: Let η be an arbitrary positive number and $0 < \varepsilon < \eta$:

$$E(f_T \mid \mathcal{H}_{0,T}) = E\left(\frac{f_T}{f_T^*} f_T^* \mid \mathcal{H}_{0,T}\right) \\ = I_{A_T^\varepsilon} E\left(\frac{f_T}{f_T^*} f_T^* \mid \mathcal{H}_{0,T}\right) + I_{(A_T^\varepsilon)^c} E\left(\frac{f_T}{f_T^*} f_T^* \mid \mathcal{H}_{0,T}\right).$$

Under the assumptions of the lemma,

$$I_{A_T^\varepsilon}(1-\varepsilon)E(f_T^* \mid \mathcal{H}_{0,T}) + I_{(A_T^\varepsilon)^c} E(f_T \mid \mathcal{H}_{0,T}) \\ \leq E(f_T \mid \mathcal{H}_{0,T}) \\ \leq I_{A_T^\varepsilon}(1+\varepsilon)E(f_T^* \mid \mathcal{H}_{0,T}) + I_{(A_T^\varepsilon)^c} E(f_T \mid \mathcal{H}_{0,T}).$$

To simplify the notation, we denote $\frac{E(f_T|\mathcal{H}_{0,T})}{E(f_T^*|\mathcal{H}_{0,T})}$ by X_T; then we get

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

We have

$$P(|X_T - 1| < \eta) \\ = P(1-\eta < X_T < 1+\eta) \\ \geq P\big[\big(I_{A_T^\varepsilon}(1+\varepsilon) + I_{(A_T^\varepsilon)^c} X_T < 1+\eta\big) \\ \cap \big(I_{A_T^\varepsilon}(1-\varepsilon) + I_{(A_T^\varepsilon)^c} X_T > 1-\eta\big)\big] \\ = P(A_T^\varepsilon) + P\big((A_T^\varepsilon)^c\big) P(1-\eta < X_T < 1+\eta) \\ \geq P(A_T^\varepsilon) \to 1,$$

where the last equality follows from the law of total probability. Hence $X_T \xrightarrow{p} 1$. *Q.E.D.*

LEMMA C.4: *Let $x_{i,T}$ be $\mathcal{H}_{i,T}$-measurable random variables and let $\Delta_{i,T} = E(x_{i,T} \mid \mathcal{H}_{i-1,T})$. Assume there are bounds $C_T \to 0$ and $D_T \to 0$ $\mathcal{H}_{0,T}$-measurable and independent of β such that*

$$\left| \sum_{i=1}^{B_N} \Delta_{i,T} \right| \leq C_T \tag{C.7}$$

and

$$\sum_{i=1}^{B_N} \Delta_{i,T}^2 \leq D_T. \tag{C.8}$$

Then

$$E\left[\prod_{i=1}^{B_N}(1+x_{i,T}) \,\middle|\, \mathcal{H}_{0,T}\right] \xrightarrow{P} 1. \tag{C.9}$$

PROOF: Using a Taylor expansion, we see that Conditions (C.7) and (C.8) imply that

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

or, more precisely,

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

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

Using iterated expectations and the definition of $\Delta_{i,T}$, we obtain

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

Hence we have

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

or, equivalently,

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

As $P(A_T^\varepsilon) \to 1$, it follows that $|E[\prod_{i=1}^{B_N}(1+x_{i,T}) \mid \mathcal{H}_{0,T}] - 1| \xrightarrow{P} 0.$ *Q.E.D.*

LEMMA C.5: *Let $a_1, a_2, \ldots, a_N$ be a sequence of numbers for some integer $N \geq 1$. Then*

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

PROOF: 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 $\sum_{i=1}^{N} p_i^l = 1/N^{l-1}$. *Q.E.D.*

C.3. *Step 1 of the Proof of Theorem 4.1*

Using a Taylor expansion, we obtain

$$\begin{aligned}
&\left| \sum_{t=1}^{T} \left(l_t\left(\theta_T + \eta_t / T^{1/4}\right) - l_t(\theta_T) \right) - \mathrm{TE}_T \right| \\
&\quad \leq \sum_{t=1}^{T} \left\| l_t^{(5)}(\theta_T) \right\| \cdot M^5 \cdot \frac{1}{T \sqrt[4]{T}} = \frac{1}{T} \sum_{t=1}^{T} \left\| l_t^{(5)}(\theta_T) \right\| \cdot M^5 \cdot \frac{1}{\sqrt[4]{T}} \\
&\quad = o_p(1)
\end{aligned}$$

because

$$\frac{1}{T} \sum_{t=1}^{T} E \left\| l_t^{(5)}(\theta_T) \right\| \leq \sup_{\theta \in \mathcal{N}} E\left(\left\| l_t^{(5)}(\theta) \right\| \right) < \infty$$

by Assumption 2. Then (C.4) follows directly from Lemma C.3.

C.4. *Step 2 of the Proof of Theorem 4.1*

Our primary objective is the computation of the conditional expectation $E(\exp(\mathrm{TE}_T) \mid \mathcal{H}_{0,T})$, or equivalently, to show that, for some $\mathcal{H}_{0,T}$-measurable Z_T, $\lim E(\exp(\mathrm{TE}_T - Z_T) \mid \mathcal{H}_{0,T}) = 1$. Let us assume that the ith block lies between $T_{i-1} + 1$ and T_i. Define

$$L_i^{(1)} = \sum_{t=T_{i-1}+1}^{T_i} l_t^{(1)}(\eta_t),$$

$$L_i^{(2)} = \sum_{t=T_{i-1}+1}^{T_i} l_t^{(2)}(\eta_t, \eta_t),$$

$$L_i^{(3)} = \sum_{t=T_{i-1}+1}^{T_i} l_t^{(3)}(\eta_t, \eta_t, \eta_t),$$

$$L_i^{(4)} = \sum_{t=T_{i-1}+1}^{T_i} l_t^{(4)}(\eta_t, \eta_t, \eta_t, \eta_t).$$

Hence we can write (C.5) as

$$\lim E\Biggl(\exp\Biggl(\sum_{i=1}^{B_N}\biggl(\frac{1}{\sqrt[4]{T}}L_i^{(1)} + \frac{1}{2\sqrt{T}}L_i^{(2)} + \frac{1}{6\sqrt[4]{T^3}}L_i^{(3)} + \frac{1}{24T}L_i^{(4)}\biggr) - Z_T\Biggr) \Bigm| \mathcal{H}_{0,T}\Biggr) = 1.$$

Let R_i be a sequence of $\mathcal{H}_{i,T}$-measurable random variables. Then define $R_{B_N+1} = R_0 = 0$. Let the function $\Gamma_i(R)$ be defined as

$$\Gamma_i(R) = E(R_{i+1} \mid \mathcal{H}_{i,T}) - E(R_{i+1} \mid \mathcal{H}_{0,T}) - E(R_i \mid \mathcal{H}_{i-1,T}) \tag{C.10}$$

(i.e., Γ_i is an $\mathcal{H}_{i,T}$-measurable random variable defined by the arguments R_i). Define

$$\begin{aligned}
M_i^{(1)} &= L_i^{(1)} + \Gamma_i\bigl(L^{(1)}\bigr),\\
M_i^{(2)} &= L_i^{(2)} + \Gamma_i\bigl(L^{(2)}\bigr) + \Gamma_i\bigl(\bigl(M^{(1)}\bigr)^2\bigr),\\
M_i^{(3)} &= L_i^{(3)} + \Gamma_i\bigl(L^{(3)}\bigr) + \Gamma_i\bigl(\bigl(M^{(1)}\bigr)^3\bigr) + \Gamma_i\bigl(3M^{(1)}M^{(2)}\bigr),\\
M_i^{(4)} &= L_i^{(4)} + \Gamma_i\bigl(L^{(4)}\bigr) + \Gamma_i\bigl(\bigl(M^{(1)}\bigr)^4\bigr) + \Gamma_i\bigl(3\bigl(M^{(2)}\bigr)^2\bigr)\\
&\quad + \Gamma_i\bigl(6\bigl(M^{(1)}\bigr)^2M^{(2)}\bigr) + \Gamma_i\bigl(4M^{(1)}M^{(3)}\bigr),
\end{aligned}$$

where we use the convention that $L_{T+1}^{(k)} = L_0^{(k)} = 0$, $\mathcal{H}_{-1,T} = \mathcal{H}_{0,T}$. For each sequence R_i of $\mathcal{H}_{i,T}$-measurable random variables, we have

$$\sum_{i=1}^{B_N}\Gamma_i(R) + \sum_{i=1}^{B_N}E(R_i \mid \mathcal{H}_{0,T}) = 0. \tag{C.11}$$

Now, we define

$$Z_T = \frac{1}{\sqrt[4]{T}}\sum_{i=1}^{B_N}E\bigl(L_i^{(1)} \mid \mathcal{H}_{0,T}\bigr) \tag{C.12}$$

$$\qquad + \frac{1}{2\sqrt{T}}\sum_{i=1}^{B_N}E\bigl(\bigl(M_i^{(1)}\bigr)^2 + L_i^{(2)} \mid \mathcal{H}_{0,T}\bigr) \tag{C.13}$$

$$(C.14)\qquad + \frac{1}{6\sqrt[4]{T^3}} \sum_{i=1}^{B_N} E\big((M_i^{(1)})^3 + L_i^{(3)} + 3M_i^{(1)}M_i^{(2)} \mid \mathcal{H}_{0,T}\big)$$

$$(C.15)\qquad + \frac{1}{24T} \sum_{i=1}^{B_N} E\big((M_i^{(1)})^4 + L_i^{(4)} + 4M_i^{(1)}M_i^{(3)} + 6(M_i^{(1)})^2 M_i^{(2)} + 3(M_i^{(2)})^2 \mid \mathcal{H}_{0,T}\big).$$

So, using the definition of the M's and (C.11), we have

$$\sum_{i=1}^{B_N}\left(\frac{1}{\sqrt[4]{T}}M_i^{(1)} + \frac{1}{2\sqrt{T}}M_i^{(2)} + \frac{1}{6\sqrt[4]{T^3}}M_i^{(3)} + \frac{1}{24T}M_i^{(4)}\right) = \sum_{i=1}^{B_N}\left(\frac{1}{\sqrt[4]{T}}L_i^{(1)} + \frac{1}{2\sqrt{T}}L_i^{(2)} + \frac{1}{6\sqrt[4]{T^3}}L_i^{(3)} + \frac{1}{24T}L_i^{(4)}\right) - Z_T.$$

Hence, to prove (C.5), it is sufficient to show that

$$(C.16)\qquad \lim E\left(\exp\left(\sum_{i=1}^{B_N}\left(\frac{1}{\sqrt[4]{T}}M_i^{(1)} + \frac{1}{2\sqrt{T}}M_i^{(2)} + \frac{1}{6\sqrt[4]{T^3}}M_i^{(3)} + \frac{1}{24T}M_i^{(4)}\right)\right) \Bigm| \mathcal{H}_{0,T}\right) = 1.$$

Before establishing (C.16), we need some preliminary results.

LEMMA C.6: *Each $M_i^{(k)}$, $k = 1, \dots, 4$ can be written as a finite (the number of terms is independent of T) sum of expressions of the form*

$$(C.17)\qquad \sum_{t_1} \cdots \sum_{t_m} \big(l_{t_1}^{(k_1)} \otimes \cdots \otimes l_{t_m}^{(k_m)}\big)(\alpha_{t_1,\dots,t_m}),$$

where $\alpha_{t_1,\dots,t_m}$ is a (random) multivariate form measurable with respect to the σ-algebra generated by $\vartheta_{T_{i-1}}, \vartheta_{T_{i-1}+1}, \dots, \vartheta_{T_i}$ denoted $\mathcal{A}_i$ and

$$(C.18)\qquad \|\alpha_{t_1,\dots,t_m}\| \le M^{\sum k_i},$$

where M and ϑ_t are defined in Assumption 3. Moreover,

$$(C.19)\qquad m \le k \quad \text{and} \quad \sum k_i = k.$$

Note that, for reasons of simplicity, we suppressed the range of summation of $t_1, \dots, t_m$; t_j *varies between* $T_{i-1+s}+1$ *and* T_{i+s}, *with* s *equal to either* 0, 1, 2, 3, *or* 4, *so each summation extends over* B_L *terms.*

DEFINITION C.7: Let us define the "order" of the expression (C.17) to be k.

REMARK 1: $l_{t_r}^{(k_r)}$ $(r = 1, \dots, 4)$ do not need to be different. This way, we can obtain power functions of the $l_{t_r}^{(k_r)}$.

PROOF OF LEMMA C.6: First of all, let us observe that all the $L_i^{(k)}$, $L_{i-1}^{(k)}$ can be written in the form of (C.17). Let us prove the lemma by induction with respect to k. The lemma holds for $k = 1$ by Assumption 3. Suppose now that we have shown our lemma for all $k \le p$. Then we have to show that it holds true for $k = p + 1$. For this purpose, observe that $M_i^{(p+1)}$ is constructed from the $M_i^{(p)}$, $M_i^{(p-1)}$ by the following operations:

- multiplying $M_i^{(p)}, M_i^{(p-1)}, \dots, L_i^{(p+1)}, L_{i-1}^{(p)}, \dots$ so that the order of the resulting tensor remains equal to $p + 1$
- adding or subtracting these products
- taking conditional expectation with respect to $\mathcal{H}_{i,T}$ or $\mathcal{H}_{i-1,T}$.

Therefore, we only need to show that the class of processes described by sums of terms of the structure of (C.17) is closed under these three operations. So let us begin with the first point. It is sufficient to show that the product of two sums of terms given by (C.17) of orders k and h gives us a sum of terms (C.17) with order $k + h$ (for products with more than two factors, simply iterate the procedure). Using the distributive law, we simply have to show that a product of two terms, say

$$V_1 = \sum_{t_1} \cdots \sum_{t_m} \big(l_{t_1}^{(k_1)} \otimes \cdots \otimes l_{t_m}^{(k_m)}\big)(\alpha_{t_1,\dots,t_m})$$

and

$$V_2 = \sum_{s_1} \cdots \sum_{s_q} \big(l_{s_1}^{(h_1)} \otimes \cdots \otimes l_{s_q}^{(h_q)}\big)(\beta_{s_1,\dots,s_q}),$$

is again of the form (C.17). But this is an immediate consequence of the distributive law and the definition of the tensor product:

$$V_1 V_2 = \sum_{t_1,t_2,\dots,t_m} \sum_{s_1,s_2,\dots,s_q} \big(l_{t_1}^{(k_1)} \otimes \cdots \otimes l_{t_m}^{(k_m)} \otimes l_{s_1}^{(h_1)} \otimes \cdots \otimes l_{s_q}^{(h_q)}\big)(\alpha_{t_1,\dots,t_m} \otimes \beta_{s_1,\dots,s_q}).$$

It follows that the order of $V_1 V_2$ equals $\sum k_j + \sum h_j = k + h$. Our class is, by construction, closed under linear operations like addition and subtraction of terms (C.17). Therefore, it remains to show that the class is closed under conditional expectations with respect to $\mathcal{H}_{i-1,T}$ and $\mathcal{H}_{i,T}$. So let us consider a term of the form

$$V = \sum_{t_1} \cdots \sum_{t_m} \big(l_{t_1}^{(k_1)} \otimes \cdots \otimes l_{t_m}^{(k_m)}\big)(\alpha_{t_1,\dots,t_m}).$$

Since $l_{t_1}^{(k_1)}, \dots, l_{t_m}^{(k_m)}$ are $\mathcal{H}_{0,T}$-measurable, we have, for $j = i, i-1$,

$$E(V \mid \mathcal{H}_{j,T}) = \sum_{t_1} \cdots \sum_{t_m} \big(l_{t_1}^{(k_1)} \otimes \cdots \otimes l_{t_m}^{(k_m)}\big) E(\alpha_{t_1,\dots,t_m} \mid \mathcal{H}_{j,T}).$$

Hence, it is sufficient to show that $E(\alpha_{t_1,\dots,t_m} \mid \mathcal{H}_{j,T})$ is measurable with respect to the σ-algebra $\mathcal{A}_i$ if it is true for $\alpha_{t_1,\dots,t_m}$. Note that $\mathcal{H}_{j,T}$ is the σ-algebra generated by $\mathcal{H}_{0,T}$ and $\sigma(\vartheta_{T_j}, \vartheta_{T_j-1}, \dots, \vartheta_1)$ (the σ-algebra generated by $(\vartheta_{T_j}, \vartheta_{T_j-1}, \dots, \vartheta_1)$) and, moreover, $\sigma(\vartheta_{T_j}, \vartheta_{T_j-1}, \dots, \vartheta_1)$ and $\mathcal{H}_{0,T}$ are independent σ-algebras, hence

$$E(\alpha_{t_1,\dots,t_m} \mid \mathcal{H}_{j,T}) = E\big(\alpha_{t_1,\dots,t_m} \mid \sigma(\vartheta_{T_j}, \vartheta_{T_j-1}, \dots, \vartheta_1)\big)$$

because $\alpha_{t_1,\dots,t_m}$ is $\mathcal{A}_i$-measurable and hence $\sigma(\vartheta_{T_j}, \vartheta_{T_j-1}, \dots, \vartheta_1)$ measurable. Moreover, we did assume that the process ϑ_t is Markov. Hence, $E(\alpha_{t_1,\dots,t_m} \mid \sigma(\vartheta_{T_j}, \vartheta_{T_j-1}, \dots, \vartheta_1)) = E(\alpha_{t_1,\dots,t_m} \mid \vartheta_{T_j})$ and therefore, $E(\alpha_{t_1,\dots,t_m} \mid \mathcal{H}_{j,T})$ is $\mathcal{A}_i$-measurable. This completes the proof of Lemma C.6. *Q.E.D.*

DEFINITION C.8: Define Λ_i by

$$\Lambda_i = \max_{k=1,2,3,4} \sum_{T_{i-1}+1}^{T_i} \big\| l_t^{(k)} \big\|.$$

It follows from Lemma C.6 that $M_i^{(k)}$ can be written as a finite sum of K elements of the form (C.17) so that

$$M_i^{(k)} = \sum_{p=1}^{K} \sum_{t_1(p)} \cdots \sum_{t_{m(p)}(p)} \big(l_{t_1(p)}^{(k_1(p))} \otimes \cdots \otimes l_{t_{m(p)}(p)}^{(k_m(p))}\big)(\alpha_{t_1(p),\dots,t_{m(p)}(p)}).$$

Then we have

$$\begin{aligned} \big|M_i^{(k)}\big| &\le \sum_{p=1}^{K} \Bigg(\sum_{t_1(p)} \big\| l_{t_1(p)}^{(k_1(p))} \big\| \cdots \sum_{t_{m(p)}(p)} \big\| l_{t_{m(p)}(p)}^{(k_m(p))} \big\| \Bigg) \|\alpha_{t_1(p),\dots,t_{m(p)}(p)}\| \\ &\le Const \sum_{p=1}^{K} \Lambda_{j_1(p)} \cdot \dots \cdot \Lambda_{j_{m(p)}(p)}. \end{aligned}$$

Hence, we can state the following corollary.

COROLLARY C.9: *There exist constants Const and K so that, for all i, k,*

$$\big|M_i^{(k)}\big| \le Const \sum_{p=1}^{K} \Lambda_{j_1(p)} \cdot \cdots \cdot \Lambda_{j_{m(p)}(p)}, \tag{C.20}$$

where, for each index p in the sum on the right hand side of inequality (C.20),

$$m(p) \le 4,$$
$$j_1(p), \dots, j_m(p) = i, i+1, \dots, or\ i+4.$$

Let us return to our goal, which is the proof of (C.16). We will establish (C.16) by showing that there exist $\mathcal{H}_{0,T}$-measurable random variables B_T and events $A_T \in \mathcal{H}_{0,T}$ so that $P(A_T) \to 1$ and $B_T \to 0$ with

$$S_i = \frac{1}{\sqrt[4]{T}} M_i^{(1)} + \frac{1}{2\sqrt{T}} M_i^{(2)} + \frac{1}{6\sqrt[4]{T^3}} M_i^{(3)} + \frac{1}{24T} M_i^{(4)}, \tag{C.21}$$

$$\sum_{i=1}^{B_N} \big| E\big(\exp(S_i) \mid \mathcal{H}_{i-1,T}\big) - 1 \big| I_{A_T} \le B_T. \tag{C.22}$$

For this to guarantee (C.16), define $x_{i,T} = (\exp(S_i) - 1) I_{A_T} I_{[B_T \le 1]}$ and apply Lemma C.4. We can set $C_T = B_T$, and since $x_{i,T} = 0$ if $B_T > 1$, and the event $[B_T > 1]$ is $\mathcal{H}_{0,T}$-measurable, we can conclude that

$$E(x_{i,T} \mid \mathcal{H}_{i-1,T}) = I_{A_T} I_{[B_T \le 1]} E\big(\big(\exp(S_i) - 1\big) \mid \mathcal{H}_{i-1,T}\big) (\equiv \Delta_{i,T})$$

and it follows from (C.22) that $\sum_{i=1}^{B_N} |\Delta_{i,T}| \le B_T$, hence $\sum_{i=1}^{B_N} \Delta_{i,T}^2 \le \sum_{i=1}^{B_N} |\Delta_{i,T}| \le B_T$, so we can set $D_T = B_T$. Then Lemma C.4 allows us to conclude that

$$E\left[\prod_{i=1}^{B_N} \big(1 + \big(\exp(S_i) - 1\big) I_{A_T} I_{[B_T \le 1]}\big) \,\Big|\, \mathcal{H}_{0,T} \right] \overset{P}{\to} 1.$$

But as A_T and $[B_T \le 1]$ are $\mathcal{H}_{0,T}$-measurable, we have

$$E\left[\prod_{i=1}^{B_N} \big(1 + \big(\exp(S_i) - 1\big) I_{A_T} I_{[B_T \le 1]}\big) \,\Big|\, \mathcal{H}_{0,T} \right]$$
$$= E\left[\prod_{i=1}^{B_N} \big(1 + \big(\exp(S_i) - 1\big)\big) \,\Big|\, \mathcal{H}_{0,T} \right] I_{A_T} I_{[B_T \le 1]} + (1 - I_{A_T} I_{[B_T \le 1]}).$$

Since $I_{A_T} I_{[B_T \le 1]} \to 1$, this implies (C.16).

Now, we focus on proving (C.22). Each $\mathcal{H}_{i-1,T} \supseteq \mathcal{H}_{0,T}$, and since $L_j^{(k)} = \sum_t l_t^{(k)}(\eta_t, \dots, \eta_t)$, where the $l_t^{(k)}$ are $\mathcal{H}_{0,T}$-measurable and the η_t are bounded, conditional expectations like $E(\exp(S_i) \mid \mathcal{H}_{i-1,T})$ exist.

We will show (C.22) in two steps:

First of all, we will show that there exist events $A_T^{(1)} \in \mathcal{H}_{0,T}$ with $P(A_T^{(1)}) \to 1$ and $\mathcal{H}_{0,T}$-measurable random variables $B_T^{(1)}$ with $B_T^{(1)} \to 0$ such that

$$\text{(C.23)} \quad \sum_{i=1}^{B_N} \Big| E\big(\exp(S_i) \mid \mathcal{H}_{i-1,T}\big) - E\left(1 + S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4 \mid \mathcal{H}_{i-1,T}\right) \Big| I_{A_T^{(1)}} \le B_T^{(1)}.$$

Second, we will show the existence of events $A_T^{(2)} \in \mathcal{H}_{0,T}$ with $P(A_T^{(2)}) \to 1$ and $\mathcal{H}_{0,T}$-measurable random variables $B_T^{(2)}$ with $B_T^{(2)} \to 0$ such that

$$\text{(C.24)} \quad \sum_{i=1}^{B_N} \left| E\left(S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4 \mid \mathcal{H}_{i-1,T}\right) \right| I_{A_T^{(2)}} \le B_T^{(2)}.$$

PROOF OF (C.23): First observe that the S_i are linear combinations of the $M_i^{(k)}$ with coefficients $1/\sqrt[4]{T^k}$. Hence we may conclude from Corollary C.9 that there exist constants K', $Const'$ so that

$$|S_i| \le \frac{1}{\sqrt[4]{T}} Const' \sum_{p=1}^{K'} \Lambda_{j_1} \cdot \dots \cdot \Lambda_{j_m},$$

where (we suppress the dependence of m and j on p in the above and all subsequent formulas for simplicity)

$$\text{(C.25)} \quad m \le 4,$$

$$\text{(C.26)} \quad j_1, \dots, j_m = i, i+1, \dots, \text{or } i+4.$$

Hence

$$|S_i|^5 \le \frac{1}{\sqrt[4]{T^5}} Const'^5 K'^4 \sum_{p=1}^{K'} (\Lambda_{j_1})^5 \cdot \dots \cdot (\Lambda_{j_m})^5$$
$$\le \frac{1}{\sqrt[4]{T^5}} Const'' \sum_{p=1}^{K'} \big((\Lambda_{j_1})^{5m} + \dots + (\Lambda_{j_m})^{5m}\big),$$

with some other constant $Const''$, where the m and j satisfy (C.25) and (C.26). The first inequality follows from Lemma C.5 and the second inequality follows from the geometric-arithmetic mean inequality, which states that

$$|a_1\cdots a_m| = \sqrt[m]{|a_1|^m\cdots|a_m|^m} \le \frac{1}{m}\big(|a_1|^m+\cdots+|a_m|^m\big). \tag{C.27}$$

Each $\Lambda_j^{(k)}$ is a maximum of sums of B_L terms of the form $\|l_t^{(k)}\|$. Hence, by Lemma C.5,

$$(\Lambda_j)^{5m} \le \max_k\left(B_L^{5m-1}\sum_{t=T_{j-1}+1}^{T_j}\big\|l_t^{(k)}\big\|^{5m}\right).$$

Since $m\le 4$, we have $5m\le 20$. Let us now define the random variable Bnd_i by

$$Bnd_i = Const''\sum_{p=1}^{K'}\left(\sum_{t=T_{j_1-1}+1}^{T_{j_1}}\big\|l_t^{(k_1)}\big\|^{5m}+\cdots+\sum_{t=T_{jm-1}+1}^{T_{jm}}\big\|l_t^{(k_m)}\big\|^{5m}\right);$$

we have

$$|S_i|^5 \le \frac{1}{\sqrt[4]{T^5}}B_L^{19}Bnd_i,$$

therefore,

$$\sum_{i=1}^{B_N}|S_i|^5 \le \frac{1}{\sqrt[4]{T^5}}B_L^{19}\sum_{i=1}^{B_N}Bnd_i. \tag{C.28}$$

Under Assumption 2, we have

$$\sup_\theta E\big\|l_t^{(k)}\big\|^{20}<\infty$$

(which implies that $\sup E\|l_t^{(k)}\|^{15}<\infty$, $\sup E\|l_t^{(k)}\|^{10}<\infty$, $\sup E\|l_t^{(k)}\|^{5}<\infty$, too). Assume

$$\frac{B_L^{76}}{T}\to 0. \tag{C.29}$$

Then we have

$$E\left(\frac{1}{\sqrt[4]{T^5}}B_L^{19}\sum_{i=1}^{B_N}Bnd_i\right) = \frac{B_L^{19}}{\sqrt[4]{T}}\frac{1}{T}E\left(\sum_{i=1}^{B_N}Bnd_i\right).$$

As $B_L^{19}/\sqrt[4]{T}$ converges to zero and $\frac{1}{T}E(\sum_{i=1}^{B_N} Bnd_i) \leq \frac{1}{T}B_N \cdot \max E(Bnd_i) \leq \frac{1}{T}B_N \cdot Const'' \cdot K \cdot B_L \cdot 4 \cdot \sup\{\sup E\|l_t^{(k)}\|^{20}, \sup E\|l_t^{(k)}\|^{15}, \sup E\|l_t^{(k)}\|^{10}, \sup E\|l_t^{(k)}\|^5\}$ remains bounded, we have

$$E\left(\frac{1}{\sqrt[4]{T^5}} B_L^{19} \sum_{i=1}^{B_N} Bnd_i\right) \to 0. \tag{C.30}$$

So, let us now come back to our original task, namely, to prove (C.23). Define, for an arbitrary $\varepsilon > 0$,

$$A_T^{(1)} = \left[\frac{1}{\sqrt[4]{T^5}} B_L^{19} \sum_{i=1}^{B_N} Bnd_i \leq \varepsilon\right].$$

Then (C.30) implies that

$$P(A_T^{(1)}) \to 1.$$

To show (C.23), it is sufficient to dominate

$$\sum_{i=1}^{B_N} \left| E(\exp(S_i) \mid \mathcal{H}_{i-1,T}) - E\left(1 + S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4 \mid \mathcal{H}_{i-1,T}\right)\right| I_{A_T^{(1)}}.$$

As $A_T^{(1)} \in \mathcal{H}_{0,T}$, $A_T^{(1)} \in \mathcal{H}_{i-1,T}$, too. Hence

$$\begin{aligned}
&\left| E(\exp(S_i) \mid \mathcal{H}_{i-1,T}) - E\left(1 + S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4 \mid \mathcal{H}_{i-1,T}\right)\right| I_{A_T^{(1)}} \\
&\quad \leq E\left(\left|\exp(S_i) - \left(1 + S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4\right)\right| \mid \mathcal{H}_{i-1,T}\right) I_{A_T^{(1)}} \\
&\quad = E\left(\left|\exp(S_i) - \left(1 + S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4\right)\right| I_{A_T^{(1)}} \mid \mathcal{H}_{i-1,T}\right).
\end{aligned}$$

Using a Taylor expansion and the fact that the fifth derivative of the exponential is the exponential itself, we see that

$$\begin{aligned}
&\left|\exp(S_i) - \left(1 + S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4\right)\right| I_{A_T^{(1)}} \\
&\quad \leq \frac{1}{120}|S_i|^5 \exp(|S_i|) I_{A_T^{(1)}}.
\end{aligned}$$

Hence, in order to show (C.23), it is sufficient to dominate

$$E\sum_{i=1}^{B_N}|S_i|^5\exp(|S_i|)I_{A_T^{(1)}}.$$

Inequality (C.28), however, implies that

$$A_T^{(1)}\subseteq\left[\sum_{i=1}^{B_N}|S_i|^5\leq\varepsilon\right],$$

hence

$$\begin{aligned}\sum_{i=1}^{B_N}|S_i|^5\exp(|S_i|)I_{A_T^{(1)}}&=\left(\sum_{i=1}^{B_N}|S_i|^5\exp(|S_i|)I_{[\sum_{i=1}^{B_N}|S_i|^5\leq\varepsilon]}I_{A_T^{(1)}}\right)\\&\leq\sum_{i=1}^{B_N}|S_i|^5\exp(\sqrt[5]{\varepsilon}).\end{aligned}$$

Therefore, we only have to dominate $\sum_{i=1}^{B_N}|S_i|^5$. This can be accomplished by setting

$$B_T^{(1)}=\frac{1}{\sqrt[4]{T^5}}B_L^{19}\sum_{i=1}^{B_N}Bnd_i.$$

By construction, $B_T^{(1)}$ is $\mathcal{H}_{0,T}$-measurable. Equation (C.28) demonstrates that it dominates $\sum_{i=1}^{B_N}|S_i|^5$. Equation (C.30) shows that $B_T^{(1)}$ converges to zero, which implies (C.23). *Q.E.D.*

PROOF OF (C.24): Define $A_T^{(2)}$ to be trivial (i.e., the whole space).

From (C.21), we can see that S_i^r, $r=1,2,3,4$ is (independently of i) a finite sum of products of normalizing factors $1/\sqrt[4]{T}, 1/\sqrt{T}, 1/\sqrt[4]{T^3}, 1/T, \ldots$. and up to four terms, which must be chosen among $M_i^{(1)}$, $M_i^{(2)}$, $M_i^{(3)}$, $M_i^{(4)}$. So we have, for some constant U,

$$S_i+\frac{1}{2}S_i^2+\frac{1}{6}S_i^3+\frac{1}{24}S_i^4=\sum_{p=1}^{U}\frac{1}{(\sqrt[4]{T})^{q(p)}}M_i^{(r_1(p))}\ldots M_i^{(r_B(p))}, \tag{C.31}$$

where $B = B(p) \leq 4$. Now split up the sum on the right hand side of (C.31) in two parts, depending on whether the normalizing factor is smaller than $\frac{1}{(\sqrt[4]{T})^5}$ or not. Define

$$D_{1,i} = \sum_{p\in\{1,\dots,U\},q(p)\geq 5} \frac{1}{(\sqrt[4]{T})^{q(p)}} M_i^{(r_1(p))} \dots M_i^{(r_B(p))} \tag{C.32}$$

and

$$D_{2,i} = \sum_{p\in\{1,\dots,U\},q(p)\leq 4} \frac{1}{(\sqrt[4]{T})^{q(p)}} M_i^{(r_1(p))} \dots M_i^{(r_B(p))}. \tag{C.33}$$

Obviously, $D_{1,i} + D_{2,i} = S_i + \frac{1}{2}S_i^2 + \frac{1}{6}S_i^3 + \frac{1}{24}S_i^4$. So we can prove (C.24) by constructing $\mathcal{H}_{0,T}$-measurable $B_T^{(2,1)}$ and $B_T^{(2,2)}$, both converging to 0, so that

$$\sum_{i=1}^{B_N} \left|E(D_{1,i} \mid \mathcal{H}_{i-1,T})\right| I_{A_T^{(2)}} \leq B_T^{(2,1)} \tag{C.34}$$

and

$$\sum_{i=1}^{B_N} \left|E(D_{2,i} \mid \mathcal{H}_{i-1,T})\right| I_{A_T^{(2)}} \leq B_T^{(2,2)}.$$

Let us first analyze $D_{1,i}$. It follows from Corollary C.9 that $M_i^{(k)}$ can be dominated by a finite (where the number of summands does not depend on T) sum of products of $\Lambda_{j_1} \cdots \Lambda_{j_m}$, where we have at most four factors. We can use this fact to bound the products $M_i^{(r_1(p))} \dots M_i^{(r_B(p))}$ in the right hand side of (C.32). This product is smaller than a constant times a sum of products of 16 factors Λ_j. We constructed $D_{1,i}$ so that, for each term, the normalizing factor is smaller than or equal to $1/\sqrt[4]{T^5}$. Since the number of summands is smaller than U (and therefore uniformly bounded), we can conclude that

$$|D_{1,i}| \leq \left(1/\sqrt[4]{T^5}\right) Const' \sum_{p=1}^{K'} \Lambda_{j_1} \cdots \Lambda_{j_m},$$

where $m = m(p) \leq 16$ and $j_1 = j_1(p), \dots, j_m = j_m(p) = i, i+1, \dots,$ or $i+4$. Now define

$$b_i = \left(1/\sqrt[4]{T^5}\right) Const' \sum_{p=1}^{K'} \Lambda_{j_1} \cdots \Lambda_{j_m}$$

and

$$B_T^{(2,1)} = \sum_{i=1}^{B_N} b_i.$$

Obviously, $B_T^{(2,1)}$ satisfies (C.34) (since b_i is $\mathcal{H}_{0,T}$-measurable, and $b_i \geq |D_{1,i}|$, $|E(D_{1,i} \mid \mathcal{H}_{i-1,T})| \leq b_i$). It remains to show that $EB_T^{(2,1)} \to 0$. To establish this result, observe that

$$\begin{aligned} Eb_i &= \left(1/\sqrt[4]{T^5}\right) Const' \sum_{p=1}^{K'} E(\Lambda_{j_1} \cdot \cdots \cdot \Lambda_{j_m}) \\ &\leq \left(1/\sqrt[4]{T^5}\right) Const'' \sum_{p=1}^{K'} \left\{ E(\Lambda_{j_1})^m + \cdots + E(\Lambda_{j_m})^m \right\} \\ &\leq \left(1/\sqrt[4]{T^5}\right) Const''' \max_{m \in \{1,2,\ldots,16\}} E(\Lambda_j)^m. \end{aligned}$$

So we need to give a bound for the moments of Λ_j. It follows from Definition C.8 and Lemma C.5 that

$$E(\Lambda_j)^m \leq B_L^m \max_k E\left\| l_t^{(k)} \right\|^m.$$

Since our assumptions imply that the moments of order 16 exist for the norms of $l_t^{(k)}$, the second factor remains bounded and we can conclude that

$$Eb_i \leq Const^{(\mathrm{IV})} \left(1/\sqrt[4]{T^5}\right) B_L^{16}.$$

Hence we have

$$EB_T^{(2,1)} = \sum_{i=1}^{B_N} Eb_i \leq Const^{(\mathrm{IV})} \left(1/\sqrt[4]{T^5}\right) B_L^{16} B_N = Const^{(\mathrm{IV})} B_L^{15} / \sqrt[4]{T},$$

which converges to zero, as $B_L^{60} = o(T)$.

It now remains to analyze $D_{2,i}$. By definition,

$$\begin{aligned} D_{2,i} &= \frac{1}{\sqrt[4]{T}} M_i^{(1)} + \frac{1}{2\sqrt{T}} \left(\left(M_i^{(1)}\right)^2 + M_2 \right) \\ &\quad + \frac{1}{6\sqrt[4]{T^3}} \left(\left(M_i^{(1)}\right)^3 + 3\left(M_i^{(1)}\right)^2 M_i^{(2)} + M_i^{(3)} \right) \\ &\quad + \frac{1}{24T} \left(\left(M_i^{(1)}\right)^4 + 4 M_i^{(1)} M_i^{(3)} \right. \\ &\quad \left. + 6\left(M_i^{(1)}\right)^2 M_i^{(2)} + 3\left(M_i^{(2)}\right)^2 + M_i^{(4)} \right). \end{aligned}$$

In order to compute $E(D_{2,i} \mid \mathcal{H}_{i-1,T})$, we will analyze $E(M_i^{(1)} \mid \mathcal{H}_{i-1,T})$, $E((M_i^{(1)})^2 + M_2 \mid \mathcal{H}_{i-1,T})$, $E((M_i^{(1)})^3 + 3(M_i^{(1)})^2 M_i^{(2)} + M_i^{(3)} \mid \mathcal{H}_{i-1,T})$, $E((M_i^{(1)})^4 + 4M_i^{(1)}M_i^{(3)} + 6(M_i^{(1)})^2 M_i^{(2)} + 3(M_i^{(2)})^2 + M_i^{(4)} \mid \mathcal{H}_{i-1,T})$ separately.

A consequence of (C.10) is that, for every $\mathcal{H}_{i,T}$-measurable sequence R_i, we have

$$E\big(R_i + \Gamma_i(R) \mid \mathcal{H}_{i-1,T}\big) = E(R_{i+1} \mid \mathcal{H}_{i-1,T}) - E(R_{i+1} \mid \mathcal{H}_{0,T}). \tag{C.35}$$

Using (C.35) and the definitions of $M_i^{(1)}$, $M_i^{(2)}$, $M_i^{(3)}$, and $M_i^{(4)}$, we have

$$E\big(M_i^{(1)} \mid \mathcal{H}_{i-1,T}\big) = E\big(L_{i+1}^{(1)} \mid \mathcal{H}_{i-1,T}\big) - E\big(L_{i+1}^{(1)} \mid \mathcal{H}_{0,T}\big), \tag{C.36}$$

$$\begin{aligned} &E\big(\big(M_i^{(1)}\big)^2 + M_i^{(2)} \mid \mathcal{H}_{i-1,T}\big) \\ &\quad = E\big(L_{i+1}^{(2)} \mid \mathcal{H}_{i-1,T}\big) - E\big(L_{i+1}^{(2)} \mid \mathcal{H}_{0,T}\big) \\ &\qquad + E\big(\big(M_{i+1}^{(1)}\big)^2 \mid \mathcal{H}_{i-1,T}\big) - E\big(\big(M_{i+1}^{(1)}\big)^2 \mid \mathcal{H}_{0,T}\big), \end{aligned} \tag{C.37}$$

$$\begin{aligned} &E\big(\big(M_i^{(1)}\big)^3 + 3\big(M_i^{(1)}\big)^2 M_i^{(2)} + M_i^{(3)} \mid \mathcal{H}_{i-1,T}\big) \\ &\quad = E\big(L_{i+1}^{(3)} \mid \mathcal{H}_{i-1,T}\big) - E\big(L_{i+1}^{(3)} \mid \mathcal{H}_{0,T}\big) \\ &\qquad + 3\big(E\big(\big(M_{i+1}^{(1)}\big)^2 M_{i+1}^{(2)} \mid \mathcal{H}_{i-1,T}\big) - E\big(\big(M_{i+1}^{(1)}\big)^2 M_{i+1}^{(2)} \mid \mathcal{H}_{0,T}\big)\big) \\ &\qquad + E\big(\big(M_{i+1}^{(1)}\big)^3 \mid \mathcal{H}_{i-1,T}\big) - E\big(\big(M_{i+1}^{(1)}\big)^3 \mid \mathcal{H}_{0,T}\big), \end{aligned} \tag{C.38}$$

$$\begin{aligned} &E\big(\big(M_i^{(1)}\big)^4 + 4M_i^{(1)}M_i^{(3)} + 6\big(M_i^{(1)}\big)^2 M_i^{(2)} + 3\big(M_i^{(2)}\big)^2 + M_i^{(4)} \mid \mathcal{H}_{i-1,T}\big) \\ &\quad = E\big(L_{i+1}^{(4)} \mid \mathcal{H}_{i-1,T}\big) - E\big(L_{i+1}^{(4)} \mid \mathcal{H}_{0,T}\big) \\ &\qquad + E\big(4M_{i+1}^{(1)}M_{i+1}^{(3)} \mid \mathcal{H}_{i-1,T}\big) - E\big(4M_{i+1}^{(1)}M_{i+1}^{(3)} \mid \mathcal{H}_{0,T}\big) \\ &\qquad + E\big(6\big(M_{i+1}^{(1)}\big)^2 M_{i+1}^{(2)} \mid \mathcal{H}_{i-1,T}\big) - E\big(6\big(M_{i+1}^{(1)}\big)^2 M_{i+1}^{(2)} \mid \mathcal{H}_{0,T}\big) \\ &\qquad + E\big(3\big(M_{i+1}^{(2)}\big)^2 \mid \mathcal{H}_{i-1,T}\big) - E\big(3\big(M_{i+1}^{(2)}\big)^2 \mid \mathcal{H}_{0,T}\big) \\ &\qquad + E\big(\big(M_{i+1}^{(1)}\big)^4 \mid \mathcal{H}_{i-1,T}\big) - E\big(\big(M_{i+1}^{(1)}\big)^4 \mid \mathcal{H}_{0,T}\big). \end{aligned} \tag{C.39}$$

We observe a pattern: all our expressions are of the form $E(R_i \mid \mathcal{H}_{i-1,T}) - E(R_i \mid \mathcal{H}_{0,T})$, where the R_i are products consisting of at most four factors from the $L_{i+1}^{(k)}$, $M_{i+1}^{(k)}$, $k = 1, \dots, 4$. This motivates the following definition.

DEFINITION C.10: A sequence R_i is called a "nice" sequence if each of the R_i can be written as a product of at most four factors from $L_{i+1}^{(k)}$, $M_{i+1}^{(k)}$, $k = 1, \dots, 4$.

REMARK 2: It will be important that we use products with factors $L_{i+1}^{(k)}$, $M_{i+1}^{(k)}$ for the construction of R_i; the index $i+1$ is essential.

LEMMA C.11: *For all nice sequences R_i, there exist $\mathcal{H}_{0,T}$-measurable random variables $B_T^{(R)}$ so that $B_T^{(R)} \to 0$ and*

$$\sum_{i=1}^{B_N} \big|E(R_i \mid \mathcal{H}_{i-1,T}) - E(R_i \mid \mathcal{H}_{0,T})\big| \le B_T^{(R)}.$$

Then, (C.24) is an immediate consequence of Lemma C.11. It suffices to take as $B_T^{(2,2)}$ the sum of all $B_T^{(R)}$ corresponding to the terms on the right hand side of (C.36), (C.37), (C.38), and (C.39). This completes the proof of Step 2 of Theorem 4.1. *Q.E.D.*

PROOF OF LEMMA C.11: Since the sequence R_i is nice, we can write it as a product of at most four factors from $L_{i+1}^{(k)}, M_{i+1}^{(k)}, k = 1, \dots, 4$. We will now apply Lemma C.6. Trivially, the lemma also applies to $L_{i+1}^{(k)}$ (simply use only one factor in the formula of the lemma). Furthermore, the reader should take notice that here we evaluate the $(i+1)$th terms instead of the ith. Then, Lemma C.6 guarantees that each $M_{i+1}^{(k)}, L_{i+1}^{(k)}, k = 1, \dots, 4$ is the sum of expressions which can be written as a finite (the number of terms is independent of T) sum of expressions of the form

$$\sum_{t_1} \cdots \sum_{t_m} \big(l_{t_1}^{(k_1)} \otimes \cdots \otimes l_{t_m}^{(k_m)}\big)(\alpha_{t_1,\dots,t_m}), \tag{C.40}$$

where $\alpha_{t_1,\dots,t_m}$ is $\mathcal{A}_{i+1}$-measurable. Since every factor making up R_i can be written as a sum of these terms, we can use the distributive law and write R_i as a finite sum of products consisting of up to four factors of the form (C.40). So we have

$$R_i = \sum_{j=1}^{K} F_{1,i}^{(j)} F_{2,i}^{(j)} F_{3,i}^{(j)} F_{4,i}^{(j)},$$

where each $F_{\cdot,i}^{(j)}$ is either 1 (if we have fewer than four factors) or has the form (C.40). Since K is finite, we can prove Lemma C.11 if we can show that, for all j, there exists an $\mathcal{H}_{0,T}$-measurable $b_T^{(j)}$ so that

$$\sum_{i=1}^{B_N} \big|E\big(F_{1,i}^{(j)} F_{2,i}^{(j)} F_{3,i}^{(j)} F_{4,i}^{(j)} \mid \mathcal{H}_{i-1,T}\big) - E\big(F_{1,i}^{(j)} F_{2,i}^{(j)} F_{3,i}^{(j)} F_{4,i}^{(j)} \mid \mathcal{H}_{0,T}\big)\big| \le b_T^{(j)} \tag{C.41}$$

and $b_T^{(j)} \to 0$.

Since $F_{\cdot,i}^{(j)}$ has the form (C.40), we can apply the distributive law and use the tensor notation to write the product as

$$F_{1,i}^{(j)}F_{2,i}^{(j)}F_{3,i}^{(j)}F_{4,i}^{(j)} = \sum_{t_1}\cdots\sum_{t_p}\big(l_{t_1}^{(k_1)}\otimes\cdots\otimes l_{t_p}^{(k_p)}\big)(\beta_{t_1,\dots,t_p}).$$

Since we compute a product of at most four factors, and each is constructed from a tensor product with order of 4 factors, we can conclude that $\sum k_j \le 16$. We suppressed the range of summation for the t_k; t_k varies between $T_{i+s}+1$ and T_{i+1+s}, with s equal to either $0, 1, 2, 3$, or 4, so that each summation extends over B_L terms. $\beta_{t_1,\dots,t_p}$ is now the tensor product of the arguments $\alpha_{t_1,\dots,t_m}$ for each of the $F_{1,i}^{(j)}, F_{2,i}^{(j)}, \dots$. The $\alpha_{t_1,\dots,t_m}$ (and the analogous expressions for $F_{2,i}^{(j)}$, $F_{3,i}^{(j)}$, $F_{4,i}^{(j)}$) are bounded by (C.18) and $\mathcal{A}_{i+1}$-measurable. Obviously, both properties carry over to $\beta_{t_1,\dots,t_p}$. It will be bounded by $M^p \le (M+1)^{16}$. So we have

$$\|\beta_{t_1,\dots,t_p}\| \le (M+1)^{16}.$$

Moreover $\beta_{t_1,\dots,t_p}$ are $\mathcal{A}_{i+1}$-measurable.

Now we will exploit Assumption 3. We define

$$b_T^{(j)} = Const(1+M)^{16}\sum_{i=1}^{B_N}\bigg(\sum_{t_1}\cdots\sum_{t_p}\big(\|l_{t_1}^{(k_1)}\|\cdots\|l_{t_p}^{(k_p)}\|\big)\bigg)\lambda^{B_L}, \tag{C.42}$$

where *Const* is a constant which we will determine later on. We show that this $b_T^{(j)}$ satisfies (C.41).

First, observe that

$$\begin{aligned}
&E\big(F_{1,i}^{(j)}F_{2,i}^{(j)}F_{3,i}^{(j)}F_{4,i}^{(j)} \mid \mathcal{H}_{i-1,T}\big) - E\big(F_{1,i}^{(j)}F_{2,i}^{(j)}F_{3,i}^{(j)}F_{4,i}^{(j)} \mid \mathcal{H}_{0,T}\big)\\
&\quad = E\bigg(\sum_{t_1}\cdots\sum_{t_p}\big(l_{t_1}^{(k_1)}\otimes\cdots\otimes l_{t_p}^{(k_p)}\big)(\beta_{t_1,\dots,t_p}) \,\Big|\, \mathcal{H}_{i-1,T}\bigg)\\
&\qquad - E\bigg(\sum_{t_1}\cdots\sum_{t_p}\big(l_{t_1}^{(k_1)}\otimes\cdots\otimes l_{t_p}^{(k_p)}\big)(\beta_{t_1,\dots,t_p}) \,\Big|\, \mathcal{H}_{0,T}\bigg)\\
&\quad = \sum_{t_1}\cdots\sum_{t_p}\big(l_{t_1}^{(k_1)}\otimes\cdots\otimes l_{t_p}^{(k_p)}\big)\\
&\qquad \times \big(E\big((\beta_{t_1,\dots,t_p}) \mid \mathcal{H}_{i-1,T}\big) - E\big((\beta_{t_1,\dots,t_p}) \mid \mathcal{H}_{0,T}\big)\big)
\end{aligned}$$

and

$$\begin{aligned}
&\Big|\sum_{t_1}\cdots\sum_{t_p}\big(l_{t_1}^{(k_1)}\otimes\cdots\otimes l_{t_p}^{(k_p)}\big)\\
&\qquad\times\big(E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{i-1,T}\big)-E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{0,T}\big)\big)\Big|\\
&\quad\le\Big\|\sum_{t_1}\cdots\sum_{t_p}\big(l_{t_1}^{(k_1)}\otimes\cdots\otimes l_{t_p}^{(k_p)}\big)\Big\|\\
&\qquad\times\big\|E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{i-1,T}\big)-E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{0,T}\big)\big\|\\
&\quad\le\sum_{t_1}\cdots\sum_{t_p}\big(\|l_{t_1}^{(k_1)}\|\cdots\|l_{t_p}^{(k_p)}\|\big)\\
&\qquad\times\big\|E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{i-1,T}\big)-E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{0,T}\big)\big\|.
\end{aligned}$$

So it suffices to show that

$$\big\|E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{i-1,T}\big)-E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{0,T}\big)\big\|\le Const(1+M)^{16}\lambda^{B_L},$$

where *Const* is a constant and M and λ are defined in Assumption 3. For a proof, let us first have a look at $\beta_{t_1,\dots,t_p}$. Tensors are elements of a finite-dimensional vector space. Let b_j be basis elements of this space. Then we can write

$$\beta_{t_1,\dots,t_p}=\sum\beta_{t_1,\dots,t_p}^{(j)}b_j,$$

with some (scalar) random variables $\beta_{t_1,\dots,t_p}^{(j)}$. We then have

$$\begin{aligned}
&\big\|E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{i-1,T}\big)-E\big((\beta_{t_1,\dots,t_p})\mid\mathcal{H}_{0,T}\big)\big\|\\
&\quad\le\Big(\sum_j\big|E\big(\big(\beta_{t_1,\dots,t_p}^{(j)}\big)\mid\mathcal{H}_{i-1,T}\big)-E\big(\big(\beta_{t_1,\dots,t_p}^{(j)}\big)\mid\mathcal{H}_{0,T}\big)\big|\Big)\Big(\max_j\|b_j\|\Big).
\end{aligned}$$

Since the sum consists only of finitely many terms, it is sufficient to show that

$$\big|E\big(\big(\beta_{t_1,\dots,t_p}^{(j)}\big)\mid\mathcal{H}_{i-1,T}\big)-E\big(\big(\beta_{t_1,\dots,t_p}^{(j)}\big)\mid\mathcal{H}_{0,T}\big)\big|\le Const(1+M)^{16}\lambda^{B_L},\tag{C.43}$$

where *Const* may be another constant than the one in (C.43). We can choose b_j so that

$$\big|\beta_{t_1,\dots,t_p}^{(j)}\big|\le(1+M)^{16}.$$

Furthermore, $\beta_{t_1,\dots,t_p}^{(j)}$ is $\mathcal{A}_{i+1}$-measurable. The result then immediately follows from the geometric ergodicity; see Assumption 3.

Now we go back to (C.42). By (C.27), we have

$$\|l_{t_1}^{(k_1)}\| \cdots \|l_{t_p}^{(k_p)}\| \leq Const(\|l_{t_1}^{(k_1)}\|^p + \cdots + \|l_{t_p}^{(k_p)}\|^p),$$

and since $p \leq 16$,

$$Eb_T^{(j)} \leq B_N \cdot Const \cdot B_L^{16} \cdot \Big(16 \max_{p\in\{1,2,\ldots,16\}} E\|l_t^{(k)}\|^p\Big) \cdot \lambda^{B_L}.$$

Note that $B_N \cdot B_L^{16} \cdot \lambda^{B_L} = TB_L^{15}\lambda^{B_L} \to 0$ when B_L satisfies (C.29). This completes the proof of Lemma C.11. *Q.E.D.*

C.5. *Step 3 of the Proof of Theorem 4.1*

In this section, we prove (C.6). We examine the terms (C.12) to (C.15) of Z_T successively.

C.5.1. *First Order Term (C.12)*

$$E(L_i^{(1)} \mid \mathcal{H}_{0,T}) = \sum_{t=T_{i-1}+1}^{T_i} l_t^{(1)} E(\eta_t) = 0$$

because $E(\eta_t) = 0$.

C.5.2. *Second Order Term (C.13)*

Recall

$$M_i^{(1)} = L_i^{(1)} + \Gamma_i(L^{(1)}).$$

By definition,

$$\Gamma_i(L^{(1)}) = E(L_{i+1}^{(1)} \mid \mathcal{H}_{i,T}) - E(L_{i+1}^{(1)} \mid \mathcal{H}_{0,T}) - E(L_i^{(1)} \mid \mathcal{H}_{i-1,T}).$$

However, $E(L_{i+1}^{(1)} \mid \mathcal{H}_{0,T}) = 0$. Hence

$$M_i^{(1)} = L_i^{(1)} - E(L_i^{(1)} \mid \mathcal{H}_{i-1,T}) + E(L_{i+1}^{(1)} \mid \mathcal{H}_{i,T})$$

and

$$M_i^{(1)2} = (L_i^{(1)} - E(L_i^{(1)} \mid \mathcal{H}_{i-1,T}))^2 + [E(L_{i+1}^{(1)} \mid \mathcal{H}_{i,T})]^2 + 2(L_i^{(1)} - E(L_i^{(1)} \mid \mathcal{H}_{i-1,T}) \cdot E(L_{i+1}^{(1)} \mid \mathcal{H}_{i,T}).$$

We decompose η_t in the following manner:

$$\eta_t = \xi_t + \alpha_t,$$
$$\xi_t = \eta_t - E(\eta_t \mid \mathcal{H}_{i-1,T}),$$
$$\alpha_t = E(\eta_t \mid \mathcal{H}_{i-1,T}).$$

Then, the conditional expectation of $M_i^{(1)2}$ is as follows:

$$\begin{aligned}(\text{C.44}) \quad E\big(M_i^{(1)2} \mid \mathcal{H}_{0,T}\big) = {} & \sum_{t,s=T_{i-1}+1}^{T_i} l_t^{(1)} \otimes l_s^{(1)} E(\xi_t \otimes \xi_s \mid \mathcal{H}_{0,T}) \\ & + \sum_{t,s=T_i+1}^{T_{i+1}} l_t^{(1)} \otimes l_s^{(1)} E(\alpha_t \otimes \alpha_s \mid \mathcal{H}_{0,T}) \\ & + 2 \sum_{t=T_{i-1}+1}^{T_i} \sum_{s=T_i+1}^{T_{i+1}} l_t^{(1)} \otimes l_s^{(1)} E(\xi_t \otimes \alpha_s \mid \mathcal{H}_{0,T}).\end{aligned}$$

In fact, with the definition of the blocks, we have

$$\sum_{i=0}^{B_N} \sum_{t,s=T_i+1}^{T_{i+1}} = \sum_{i=1}^{B_N+1} \sum_{t,s=T_{i-1}+1}^{T_i} = \sum_{i=1}^{B_N} \sum_{t,s=T_{i-1}+1}^{T_i}.$$

Thus, taking the summation of the first two terms in (C.44) over the blocks yields

$$\sum_{i=1}^{B_N} \sum_{t,s=T_{i-1}+1}^{T_i} l_t^{(1)} \otimes l_s^{(1)} \big[E(\xi_t \otimes \xi_s \mid \mathcal{H}_{0,T}) + E(\alpha_t \otimes \alpha_s \mid \mathcal{H}_{0,T})\big].$$

Meanwhile, the cross-product term has the following property:

$$\sum_{t,s=T_{i-1}+1}^{T_i} l_t^{(1)} \otimes l_s^{(1)} E(\xi_t \otimes \alpha_s \mid \mathcal{H}_{0,T}) = 0,$$

which is a direct consequence of $E(\xi_t \mid \mathcal{H}_{i-1,T}) = 0$. This, in turn, implies that the sum of the first two terms is equal to

$$\sum_{t=T_{i-1}+1}^{T_i} l_t^{(1)} \otimes l_t^{(1)} E(\eta_t \otimes \eta_t \mid \mathcal{H}_{0,T}).$$

Now we look at the summation over blocks of the third term in (C.44),

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}} l_t^{(1)}\otimes l_s^{(1)}E(\xi_t\otimes\alpha_s\mid\mathcal{H}_{0,T}).$$

Notice that $l_t^{(1)}\otimes l_s^{(1)}$ is a m.d.s.; therefore, all the terms $l_t^{(1)}\otimes l_s^{(1)}E(\xi_t\otimes\alpha_s\mid\mathcal{H}_{0,T})$ are uncorrelated. Its variance can be bounded in the following way:

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}} E\big[l_t^{(1)}\otimes l_s^{(1)}E(\xi_t\otimes\alpha_s\mid\mathcal{H}_{0,T})\big]^2\\
&\quad\le\frac{1}{T}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}}\sqrt{E\big(l_t^{(1)}\big)^4}\cdot\sqrt{E\big(l_s^{(1)}\big)^4}\cdot\sqrt{E\big[E(\xi_t\otimes\alpha_s\mid\mathcal{H}_{0,T})\big]^4}\\
&\quad\le\frac{1}{T}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}}\sqrt{E\big(l_t^{(1)}\big)^4}\cdot\sqrt{E\big(l_s^{(1)}\big)^4}\\
&\qquad\cdot\sqrt{E\big[E\big(\|\xi_t\otimes\alpha_s\|^4\mid\mathcal{H}_{0,T}\big)\big]}\\
&\quad=\frac{1}{T}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}}\sqrt{E\big(l_t^{(1)}\big)^4}\cdot\sqrt{E\big(l_s^{(1)}\big)^4}\cdot\sqrt{E\big(\|\xi_t\otimes\alpha_s\|^4\big)}\\
&\quad=\frac{1}{T}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}} Const\cdot\sqrt{E\|\xi_t\otimes\alpha_s\|^4}\\
&\quad\le\frac{1}{T}\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\sum_{s=T_i+1}^{T_{i+1}} Const\cdot\lambda^{2(s-t)}\\
&\quad=\frac{1}{T}Const\cdot\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\lambda^{-2t}\lambda^{2(T_i+1)}\sum_{u=0}^{B_L}\lambda^{2u}\\
&\quad\le\frac{1}{T}Const\cdot\sum_{i=1}^{B_N}\sum_{t=T_{i-1}+1}^{T_i}\lambda^{-2t}\lambda^{2(T_i+1)}\frac{1}{1-\lambda^2}\\
&\quad=\frac{1}{T}Const\cdot\frac{1}{1-\lambda^2}\cdot\sum_{i=1}^{B_N}\sum_{u=1}^{B_L}\lambda^{2u}
\end{aligned}$$

$$\leq \frac{1}{T} Const \cdot \frac{1}{1-\lambda^2} \cdot \sum_{i=1}^{B_N} \frac{\lambda^2}{1-\lambda^2}$$

$$= \frac{B_N}{T} Const \cdot \frac{1}{1-\lambda^2} \cdot \frac{\lambda^2}{1-\lambda^2} \to 0,$$

where the first inequality follows from the Cauchy–Schwarz inequality and the second inequality is from Jensen's inequality. Therefore, the third term in (C.44) converges in mean square to 0 and is therefore negligible. This approach will be used very often in the following context.

At the same time, we have

$$E\big(L_i^{(2)} \mid \mathcal{H}_{0,T}\big) = \sum_{t,s=T_{i-1}+1}^{T_i} l_t^{(2)} E(\eta_t \otimes \eta_s \mid \mathcal{H}_{0,T}).$$

Thus, we have

$$\text{(C.45)} \quad E\left(\frac{1}{\sqrt{T}} \sum_{i=1}^{B_N} \big(M_i^{(1)2} + L_i^{(2)}\big) \,\Big|\, \mathcal{H}_{0,T}\right)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \big(l_t^{(2)} + l_t^{(1)} \otimes l_t^{(1)}\big) E(\eta_t \otimes \eta_t)$$

$$+ \frac{2}{\sqrt{T}} \sum_{t<s} l_t^{(1)} \otimes l_s^{(1)} E(\eta_t \otimes \eta_s) + o_p(1)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mu_{2,t} + o_p(1).$$

C.5.3. *Third Order Term (C.14)*

First, the third order Bartlett identity implies that every coefficient of

$$\text{(C.46)} \quad l_t^{(3)} + \big(l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)}\big)^{(S)} + 3\big(l_t^{(1)} \otimes l_t^{(2)}\big)^{(S)}$$

is a m.d.s. To evaluate the third order term, we will need to evaluate a lot of expressions as in (C.46). In most of the cases, however, the arguments of the forms are simply η_t. To simplify our expressions, we will suppress the argument $l_t^{(1)}(\eta_t)^3$ and simply write $(l_t^{(1)})^3$. So instead of

$$\big(l_t^{(3)} + \big(l_t^{(1)} \otimes l_t^{(1)} \otimes l_t^{(1)}\big)^{(S)} + 3\big(l_t^{(1)} \otimes l_t^{(2)}\big)^{(S)}\big)(\eta_t, \eta_t, \eta_t),$$

we can write

$$l_t^{(3)} + l_t^{(1)3} + 3 l_t^{(1)} l_t^{(2)}.$$

Because of the property of third-order Bartlett identity mentioned above, these terms are uncorrelated. Moreover, Assumption 2 and the boundedness of η_t guarantee that the variances of these terms are bounded. Hence,

$$1/\sqrt[4]{T^3}E\left(\left|\sum_{t=1}^{T}\left(l_t^{(3)}+l_t^{(1)3}+3l_t^{(1)}l_t^{(2)}\right)\right|\right)\to 0,$$

since the expectation of the square of the sum will be $O(T)$. Therefore, we have

$$\text{(C.47)}\quad \frac{1}{\sqrt[4]{T^3}}\sum_{i=1}^{B_N}L_i^{(3)}=-\frac{1}{\sqrt[4]{T^3}}\sum_{t=1}^{T}l_t^{(1)3}-\frac{3}{\sqrt[4]{T^3}}\sum_{t=1}^{T}l_t^{(1)}l_t^{(2)}+r_T,$$

where $E|r_T|\to 0$.

Moreover,

$$\begin{aligned}-\sum_{t=1}^{T}l_t^{(1)3}&=-\left(\sum_{t=1}^{T}l_t^{(1)}\right)^3+3\sum_{t=1}^{T}l_t^{(1)2}\sum_{s<t}l_s^{(1)}+3\sum_{t=1}^{T}l_t^{(1)}\sum_{s,k<t}l_s^{(1)}l_k^{(1)}\\&=-\left(\sum_{i=1}^{B_N}L_i^{(1)}\right)^3+3\sum_{t=1}^{T}l_t^{(1)2}\sum_{s<t}l_s^{(1)}+3\sum_{t=1}^{T}l_t^{(1)}\sum_{s,k<t}l_s^{(1)}l_k^{(1)}.\end{aligned}$$

Again, the first order Bartlett condition guarantees that the conditional expectation of the derivatives of the log-likelihood function are m.d.s. Hence the $l_t^{(1)}\sum_{s,k<t}l_s^{(1)}l_k^{(1)}$ are (as a product of a m.d.s. with terms determined in the "past") m.d.s. Therefore, these terms are uncorrelated, and, again, our Assumption 3 guarantees that the variance of each term is bounded by $Const\cdot B_L^4$. Hence $E((1/\sqrt[4]{T^3}\sum l_t^{(1)}\sum_{s,k<t}l_s^{(1)}l_k^{(1)})^2)=O(T^{-1.5}\cdot B_N\cdot B_L^4)=O(T^{-1/2}B_L^3)$, so it converges to zero. Therefore,

$$E\left(\left|1/\sqrt[4]{T^3}\sum l_t^{(1)}\sum_{s,k<t}l_s^{(1)}l_k^{(1)}\right|\right)\to 0.$$

As a consequence, we have

$$\text{(C.48)}\quad E\left(\left|\frac{1}{\sqrt[4]{T^3}}\sum_{i=1}^{B_N}L_i^{(3)}-\left(-\frac{1}{\sqrt[4]{T^3}}\left(\sum_{i=1}^{B_N}L_i^{(1)}\right)^3+\frac{3}{\sqrt[4]{T^3}}\sum_{t=1}^{T}l_t^{(1)2}\sum_{s<t}l_s^{(1)}-\frac{3}{\sqrt[4]{T^3}}\sum_{t=1}^{T}l_t^{(1)}l_t^{(2)}\right)\right|\right)\to 0.$$

Meanwhile, we have

$$\left(\sum_{i=1}^{B_N} M_i^{(1)}\right)^3 = \sum_{i=1}^{B_N}\left(M_i^{(1)3} + 3M_i^{(1)2}\sum_{j<i} M_j^{(1)} + 3M_i^{(1)}\sum_{j,k<i} M_j^{(1)} M_k^{(1)}\right).$$

We can rearrange the terms and obtain

$$\text{(C.49)}\quad \sum_{i=1}^{B_N} M_i^{(1)3} = \left(\sum_{i=1}^{B_N} M_i^{(1)}\right)^3 - 3\sum_{i=1}^{B_N} M_i^{(1)2}\sum_{j<i} M_j^{(1)} - 3\sum_{i=1}^{B_N} M_i^{(1)}\sum_{j,k<i} M_j^{(1)} M_k^{(1)}.$$

First of all, let us analyze $M_i^{(1)}\sum_{j,k<i} M_j^{(1)} M_k^{(1)}$. We have

$$\text{(C.50)}\quad E\left(M_i^{(1)}\sum_{j,k<i} M_j^{(1)} M_k^{(1)} \mid \mathcal{H}_{0,T}\right) = E\left(E\left(M_i^{(1)} \mid \mathcal{H}_{i-1,T}\right)\sum_{j,k<i} M_j^{(1)} M_k^{(1)} \mid \mathcal{H}_{0,T}\right).$$

Using Equation (C.36), we have

$$E\left(M_i^{(1)} \mid \mathcal{H}_{i-1,T}\right) = E\left(L_{i+1}^{(1)} \mid \mathcal{H}_{i-1,T}\right) - E\left(L_{i+1}^{(1)} \mid \mathcal{H}_{0,T}\right).$$

From geometric ergodicity, we have

$$\text{(C.51)}\quad \sqrt{E\left(\left|E\left(M_i^{(1)} \mid \mathcal{H}_{i-1,T}\right)\right|^2\right)} \le C\cdot\lambda^{B_L}\cdot B_L,$$

which implies that we can derive a bound for (C.50) using the Cauchy–Schwarz inequality, namely,

$$E\left|\left(E\left(M_i^{(1)} \mid \mathcal{H}_{i-1,T}\right)\sum_{j,k<i} M_j^{(1)} M_k^{(1)} \mid \mathcal{H}_{0,T}\right)\right| \le C\cdot\lambda^{B_L}\cdot B_L\cdot\sqrt{E\left(\sum_{j,k<i} M_j^{(1)} M_k^{(1)}\right)^2} \le C\lambda^{B_L} B_L^2 B_N^2 = C\lambda^{B_L} T^2.$$

Since $B_L/\log T \to \infty$, $C\lambda^{B_L} T^2 \to 0$. Hence we have shown that

$$\text{(C.52)}\quad E\left(\left|\sum_{i=1}^{B_N} M_i^{(1)3} - \left(\left(\sum_{i=1}^{B_N} M_i^{(1)}\right)^3 - 3\sum_{i=1}^{B_N} M_i^{(1)2}\sum_{j<i} M_j^{(1)}\right)\right|\right) \to 0.$$

Moreover, from the definition of $M_i^{(1)}$, we have

$$\sum_{i=1}^{B_N} M_i^{(1)} = \sum_{i=1}^{B_N} L_i^{(1)} + \sum_{i=1}^{B_N} \Gamma_i(L^{(1)}).$$

Using Equation (C.11), we have

$$\sum_{i=1}^{B_N} \Gamma_i(L^{(1)}) = -\sum_{i=1}^{B_N} E(L_i^{(1)} \mid \mathcal{H}_{0,T}) = 0,$$

which implies that

$$\sum_{i=1}^{B_N} M_i^{(1)} = \sum_{i=1}^{B_N} L_i^{(1)}. \tag{C.53}$$

Now let us rewrite the third order term (C.14) using (C.48), (C.52), (C.53):

$$\begin{aligned}
&\frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} \left((M_i^{(1)})^3 + L_i^{(3)} + 3M_i^{(1)} M_i^{(2)} \right) \\
&= \frac{1}{\sqrt[4]{T^3}} \left(\sum_{i=1}^{B_N} \left(-3M_i^{(1)2} \sum_{j<i} M_j^{(1)} + 3M_i^{(1)} M_i^{(2)} \right) \right. \\
&\qquad \left. + 3 \sum_{t=1}^{T} \left(l_t^{(1)2} \sum_{s<t} l_s^{(1)} - l_t^{(1)} l_t^{(2)} \right) \right) + r_T \qquad \text{(C.54)} \\
&= \frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} \left(-3M_i^{(1)2} \sum_{j<i} M_j^{(1)} - 3M_i^{(1)3} + 3l_t^{(1)2} \sum_{s<t} l_s^{(1)} + 3l_t^{(1)3} \right) \qquad \text{(C.55)} \\
&\quad + \frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} \left(3M_i^{(1)} M_i^{(2)} + 3M_i^{(1)3} - 3l_t^{(1)} l_t^{(2)} - 3l_t^{(1)3} \right) + r_T, \qquad \text{(C.56)}
\end{aligned}$$

with $E|r_T| \to 0$.

The last transformation results from a simple rearrangement of the terms as well as a trivial algebraic operation. We added the terms $-3M_i^{(1)3}$ and $3l_t^{(1)3}$ in (C.55) and subtracted them in (C.56). Hence it is sufficient to show that

$$\begin{aligned}
E\Bigg(\frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} \bigg(& -3M_i^{(1)2} \sum_{j<i} M_j^{(1)} - 3M_i^{(1)3} \\
& + 3l_t^{(1)2} \sum_{s<t} l_s^{(1)} + 3l_t^{(1)3} \bigg) \,\bigg|\, \mathcal{H}_{0,T} \Bigg)
\end{aligned} \tag{C.57}$$

$$+E\left(\frac{1}{\sqrt[4]{T^3}}\sum_{i=1}^{B_N}\bigl(3M_i^{(1)}M_i^{(2)}+3M_i^{(1)3}-3l_t^{(1)}l_t^{(2)}-3l_t^{(1)3}\bigr)\,\Big|\,\mathcal{H}_{0,T}\right)\to 0. \tag{C.58}$$

The following algebraic identities describing the discrete analog to partial integration (sometimes called "partial summation") will prove to be useful. For arbitrary X_i, Y_i, we have

$$X_iY_i+X_i\sum_{j<i}Y_j+Y_i\sum_{j<i}X_j=\sum_{j\le i}X_j\sum_{j\le i}Y_j-\sum_{j\le i-1}X_j\sum_{j\le i-1}Y_j. \tag{C.59}$$

Computing the sum over the index i in (C.59) gives

$$\sum_{i=1}^{B_N}\left(X_iY_i+X_i\sum_{j<i}Y_j+Y_i\sum_{j<i}X_j\right)=\sum_{i=1}^{B_N}X_j\sum_{j=1}^{B_N}Y_j. \tag{C.60}$$

Let us first analyze (C.57). Using (C.60) twice shows that the sum in (C.57) equals

$$\begin{aligned}&\frac{1}{\sqrt[4]{T^3}}\sum_{i=1}^{B_N}E\left(3M_i^{(1)}\sum_{j<i}M_j^{(1)2}-3\sum_{i=1}^{B_N}M_i^{(1)}\sum_{i=1}^{B_N}M_i^{(1)2}\,\Big|\,\mathcal{H}_{0,T}\right)\\&\quad-\frac{1}{\sqrt[4]{T^3}}\sum_{t=1}^{T}E\left(3l_t^{(1)}\sum_{s<t}l_s^{(1)2}-3\sum_{t=1}^{T}l_t^{(1)}\sum_{t=1}^{T}l_t^{(1)2}\,\Big|\,\mathcal{H}_{0,T}\right).\end{aligned} \tag{C.61}$$

Note that

$$\left|E\left(M_i^{(1)}\sum_{j<i}M_j^{(1)2}\mid\mathcal{H}_{0,T}\right)\right|\le\sum_{j<i}\bigl|E\bigl(M_j^{(1)2}E\bigl(M_i^{(1)}\mid\mathcal{H}_{i-1,T}\bigr)\mid\mathcal{H}_{0,T}\bigr)\bigr|. \tag{C.62}$$

Then using the Cauchy–Schwarz inequality, (C.51), and the fact that $EM_j^{(1)4}\le C\cdot B_L^4$ (by our assumptions on the moments of the log-likelihood), we obtain

$$\left|E\left(M_i^{(1)}\sum_{j<i}M_j^{(1)2}\mid\mathcal{H}_{0,T}\right)\right|\le C\cdot\lambda^{B_L}\cdot B_N\cdot B_L^5. \tag{C.63}$$

So

$$E\left|1/\sqrt[4]{T^3}\sum_{i=1}^{B_N}E\left(M_i^{(1)}\sum_{j<i}M_j^{(1)2}\mid\mathcal{H}_{0,T}\right)\right|\to 0.$$

Analogously, we can see that the first order Bartlett identity implies that the random variables

$$E\left(l_t^{(1)}\sum_{s<t} l_s^{(1)2} \mid \mathcal{H}_{0,T}\right) \tag{C.64}$$

are uncorrelated (they are linear combinations of products of the coefficients of $l_t^{(1)}$ with terms dependent on "past" data, so they are even m.d.s.). Moreover, it follows from our assumptions regarding the existence of moments of $l_t^{(1)}$ as well as the mixing condition of the η_t that the variance of

$$E\left(l_t^{(1)}\sum_{s<t} l_s^{(1)2} \mid \mathcal{H}_{0,T}\right) = \sum_{s<t} E\big(l_t^{(1)} l_s^{(1)2} \mid \mathcal{H}_{0,T}\big)$$

is of order 1 (our condition on exponential mixing guarantees the convergence of the series on the right hand side). Hence,

$$E\left(\sum_{t=1}^{T} E\left(l_t^{(1)}\sum_{s<t} l_s^{(1)2} \mid \mathcal{H}_{0,T}\right)\right)^2 = O(T)$$

and therefore,

$$\frac{1}{\sqrt[4]{T^3}} E\left(\sum_{t=1}^{T} E\left(l_t^{(1)}\sum_{s<t} l_s^{(1)2} \mid \mathcal{H}_{0,T}\right)\right) \to 0.$$

Therefore, we can simplify (C.61) and conclude that the difference of the terms in (C.57) and

$$\begin{aligned} &\frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} E\left(-3\left(\sum_{i=1}^{B_N} M_i^{(1)}\right)\left(\sum_{i=1}^{B_N} M_i^{(1)2}\right)\right. \\ &\qquad \left. + 3\frac{1}{\sqrt[4]{T^3}} \sum_{t=1}^{T} l_t^{(1)} \sum_{t=1}^{T} l_t^{(1)2} \,\middle|\, \mathcal{H}_{0,T}\right) \end{aligned} \tag{C.65}$$

converges to zero in probability.

Now we will analyze (C.58). Applying again the "partial summation" (C.59) and (C.60), we can conclude that the terms from (C.58) are equal to

$$\begin{aligned} &\frac{3}{\sqrt[4]{T^3}} \left\{ \sum_{i=1}^{B_N} E\left(\left(\sum_{i=1}^{B_N} M_i^{(1)}\right)\left(\sum_{i=1}^{B_N} \big(M_i^{(2)} + M_i^{(1)2}\big)\right) \,\middle|\, \mathcal{H}_{0,T}\right)\right. \\ &\qquad - \sum_{i=1}^{B_N} E\left(\sum_{i=1}^{B_N} M_i^{(1)}\left(\sum_{j<i} \big(M_j^{(2)} + M_j^{(1)2}\big)\right) \,\middle|\, \mathcal{H}_{0,T}\right) \end{aligned}$$

$$
\begin{aligned}
&-\sum_{i=1}^{B_N} E\left(\left(\sum_{i=1}^{B_N}\left(M_i^{(2)}+M_i^{(1)2}\right)\left(\sum_{j<i} M_j^{(1)}\right)\right) \Bigm| \mathcal{H}_{0,T}\right) \\
&-E\left(\left(\sum_{t=1}^{T} l_t^{(1)}\right)\left(\sum_{t=1}^{T}\left(l_t^{(2)}+l_t^{(1)2}\right)\right) \Bigm| \mathcal{H}_{0,T}\right) \\
&+E\left(\left(\sum_{t=1}^{T} l_t^{(1)} \sum_{s<t}\left(l_s^{(2)}+l_s^{(1)2}\right)\right) \Bigm| \mathcal{H}_{0,T}\right) \\
&+E\left(\left(\sum_{t=1}^{T}\left(l_t^{(2)}+l_t^{(1)2}\right) \sum_{s<t} l_s^{(1)}\right) \Bigm| \mathcal{H}_{0,T}\right)\Bigg\}.
\end{aligned}
$$

Perfectly analogous to our analysis of (C.57), we can see that

$$
\frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} E\left(\sum_{i=1}^{B_N} M_i^{(1)}\left(\sum_{j<i}\left(M_j^{(2)}+M_j^{(1)2}\right)\right) \Bigm| \mathcal{H}_{0,T}\right)
$$

and

$$
\frac{1}{\sqrt[4]{T^3}} E\left(\sum_{t=1}^{T} l_t^{(1)} \sum_{s<t}\left(l_s^{(2)}+l_s^{(1)2}\right) \Bigm| \mathcal{H}_{0,T}\right)
$$

converge to zero. The second Bartlett identity guarantees that all coefficients of $l_s^{(2)}+l_s^{(1)2}$ are m.d.s. Hence the random variables

$$
E\left(\left(\left(l_t^{(2)}+l_t^{(1)2}\right) \sum_{s<t} l_s^{(1)}\right) \mid \mathcal{H}_{0,T}\right)
$$

are uncorrelated. Our assumptions about the existence of moments and the exponential mixing condition immediately allow to uniformly bound the variance of these expressions. Hence,

$$
\sum_{t=1}^{T} E\left(\left(\left(l_t^{(2)}+l_t^{(1)2}\right) \sum_{s<t} l_s^{(1)}\right) \mid \mathcal{H}_{0,T}\right)=O_p(\sqrt{T}),
$$

and therefore,

$$
\frac{1}{\sqrt[4]{T^3}} \sum_{t=1}^{T} E\left(\left(\left(l_t^{(2)}+l_t^{(1)2}\right) \sum_{s<t} l_s^{(1)}\right) \mid \mathcal{H}_{0,T}\right) \rightarrow 0.
$$

Immediately from the definition of $M_i^{(2)}$, we can see from (C.37) that

$$E\big(E\big(\big(M_i^{(2)} + M_i^{(1)2}\big) \mid \mathcal{H}_{i-1,T}\big)\big)^2 \le Const \cdot \lambda^{BL}.$$

Analogously to (C.62), (C.63), we can show that

$$\frac{1}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} E\left(\left(\sum_{i=1}^{B_N} \big(M_i^{(2)} + M_i^{(1)2}\big)\left(\sum_{j<i} M_j^{(1)}\right)\right) \,\Big|\, \mathcal{H}_{0,T}\right) \to 0.$$

So instead of (C.58), we only need to consider

$$\text{(C.66)} \quad \frac{3}{\sqrt[4]{T^3}} \left\{ \sum_{i=1}^{B_N} E\left(\left(\sum_{i=1}^{B_N} M_i^{(1)}\right)\left(\sum_{i=1}^{B_N} \big(M_i^{(2)} + M_i^{(1)2}\big)\right) \,\Big|\, \mathcal{H}_{0,T}\right) - E\left(\left(\sum_{t=1}^{T} l_t^{(1)}\right)\left(\sum_{t=1}^{T} \big(l_t^{(2)} + l_t^{(1)2}\big)\right) \,\Big|\, \mathcal{H}_{0,T}\right)\right\}.$$

The first terms, (C.57), have been simplified to (C.65). We have to show that the sum of (C.65) and (C.66) converges to zero.

So we have to show that

$$\frac{3}{\sqrt[4]{T^3}} \left\{ \sum_{i=1}^{B_N} E\left(\left(\sum_{i=1}^{B_N} M_i^{(1)}\right)\left(\sum_{i=1}^{B_N} \big(M_i^{(2)}\big)\right) - \left(\sum_{t=1}^{T} l_t^{(1)}\right)\left(\sum_{t=1}^{T} \big(l_t^{(2)}\big)\right) \,\Big|\, \mathcal{H}_{0,T}\right)\right\}$$

converges to zero. We have the following simplification:

$$\begin{aligned}
&\frac{3}{\sqrt[4]{T^3}} \left(\sum_{i=1}^{B_N} M_i^{(1)} \sum_{i=1}^{B_N} M_i^{(2)} - \sum_{i=1}^{B_N} L_i^{(1)} \sum_{i=1}^{B_N} L_i^{(2)}\right) \\
&\quad = \frac{3}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} L_i^{(1)} \sum_{i=1}^{B_N} \big(M_i^{(2)} - L_i^{(2)}\big) \\
&\quad = \frac{3}{\sqrt[4]{T^3}} \sum_{i=1}^{B_N} L_i^{(1)} \sum_{i=1}^{B_N} \big(\Gamma_i\big(L^{(2)}\big) + \Gamma_i\big(M^{(1)2}\big)\big),
\end{aligned}$$

where the second equality is a straight consequence of (C.53) and the third equality holds by definition.

Note that (C.11) implies $\sum_{i=1}^{B_N}(\Gamma_i(L^{(2)})+\Gamma_i(M^{(1)2}))$ is $\mathcal{H}_{0,T}$-measurable. So

$$E\left(\sum_{i=1}^{B_N} L_i^{(1)} \sum_{i=1}^{B_N}\left(\Gamma_i\left(L^{(2)}\right)+\Gamma_i\left(M^{(1)2}\right)\right) \,\middle|\, \mathcal{H}_{0,T}\right)$$
$$=\sum_{i=1}^{B_N}\left(\Gamma_i\left(L^{(2)}\right)+\Gamma_i\left(M^{(1)2}\right)\right)E\left(\sum_{i=1}^{B_N} L_i^{(1)} \,\middle|\, \mathcal{H}_{0,T}\right)$$
$$=0,$$

which is what we wanted to prove. In conclusion, (C.14) $= o_p(1)$.

C.5.4. *Fourth Order Term (C.15)*

To deal with the fourth order term, we use the following lemma.

LEMMA C.12:

$$\frac{1}{T}\sum_{i=1}^{B_N} E\left(L_i^{(4)}+\left(L_i^{(1)}\right)^4+4L_i^{(1)}L_i^{(3)}\right.$$
$$\left.+6\left(L_i^{(1)}\right)^2 L_i^{(2)}+3\left(L_i^{(2)}\right)^2 \mid \mathcal{H}_{0,T}\right)=o_p(1).$$

PROOF: Denote

$$Z_{il}=\left(L_i^{(1)}\right)^4+L_i^{(4)}+4L_i^{(1)}L_i^{(3)}+6\left(L_i^{(1)}\right)^2 L_i^{(2)}+3\left(L_i^{(2)}\right)^2. \tag{C.67}$$

In the sequel, we use the notation $\sum_t$ for $\sum_{t=T_{i-1}+1}^{T_i}$. We have

$$L_i^{(1)2}=\left(\sum_t l_t^{(1)}\right)^2=\sum_t l_t^{(1)2}+\sum_{t\neq s} l_t^{(1)}l_s^{(1)}$$

and

$$L_i^{(1)4}=\left(\sum_t l_t^{(1)}\right)^4$$
$$=\sum_t l_t^{(1)4}+4\sum_t l_t^{(1)3}\sum_{s\neq t} l_s^{(1)}+6\sum_t l_t^{(1)2}\sum_{s\neq j\neq t} l_s^{(1)}l_j^{(1)}$$
$$+3\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(1)2}+\sum_{t\neq s\neq j\neq k} l_t^{(1)}l_s^{(1)}l_j^{(1)}l_k^{(1)}.$$

Now we rewrite (C.67) as

$$
\begin{aligned}
Z_{il} = {} & \sum_t l_t^{(1)4} + 4\sum_t l_t^{(1)3}\sum_{s\neq t} l_s^{(1)} + 6\sum_t l_t^{(1)2}\sum_{s\neq j\neq t} l_s^{(1)} l_j^{(1)} \\
& + 3\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(1)2} + \sum_{t\neq s\neq j\neq k} l_t^{(1)} l_s^{(1)} l_j^{(1)} l_k^{(1)} + \sum_t l_t^{(4)} \\
& + 4\sum_t l_t^{(1)}\sum_t l_t^{(3)} \qquad \text{(C.68)} \\
& + 6\Bigl(\sum l_t^{(1)2} + \sum_{t\neq s} l_t^{(1)} l_s^{(1)}\Bigr)\sum_t l_t^{(2)} \qquad \text{(C.69)} \\
& + 3\Bigl(\sum_t l_t^{(2)2} + \sum_{s\neq t} l_t^{(2)} l_s^{(2)}\Bigr). \qquad \text{(C.70)}
\end{aligned}
$$

Moreover, we rewrite (C.68) as

$$
4\sum_t l_t^{(1)}\sum_t l_t^{(3)} = 4\sum_t l_t^{(1)} l_t^{(3)} + 4\sum_{t\neq s} l_t^{(1)} l_s^{(3)}
$$

and rewrite (C.69) as

$$
\begin{aligned}
& 6\Bigl(\sum l_t^{(1)2} + \sum_{t\neq s} l_t^{(1)} l_s^{(1)}\Bigr)\sum_t l_t^{(2)} \\
& = 6\sum_t l_t^{(1)2} l_t^{(2)} + 6\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(2)} + 6\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(1)}\sum_t l_t^{(2)} \\
& = 6\sum_t l_t^{(1)2} l_t^{(2)} + 6\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(2)} \\
& \quad + 12\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(1)} l_s^{(2)} + 6\sum_{t\neq s\neq j} l_t^{(1)} l_s^{(1)} l_j^{(2)}.
\end{aligned}
$$

After replacing (C.68) and (C.69) by their expressions and rearranging the terms, we obtain

$$
\begin{aligned}
Z_{il} = {} & \sum_t l_t^{(4)} + 6\sum_t l_t^{(2)} l_t^{(1)2} + 4\sum_t l_t^{(3)} l_t^{(1)} + 3\sum_t l_t^{(2)2} + \sum_t l_t^{(1)4} \\
& + 4\sum_t l_t^{(1)3}\sum_{s\neq t} l_s^{(1)} + 6\sum_t l_t^{(1)2}\sum_{s\neq j\neq t} l_s^{(1)} l_j^{(1)} \\
& + 3\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(1)2} + \sum_{t\neq s\neq j\neq k} l_t^{(1)} l_s^{(1)} l_j^{(1)} l_k^{(1)}
\end{aligned} \qquad \text{(C.71)}
$$

$$
\begin{aligned}
&+4\sum_{t\neq s} l_t^{(1)} l_s^{(3)} \\
&+6\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(2)} + 12\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(1)} l_s^{(2)} + 6\sum_{t\neq s\neq j} l_t^{(1)} l_s^{(1)} l_j^{(2)} \\
&+3\Big(\sum_{s\neq t} l_t^{(2)} l_s^{(2)}\Big),
\end{aligned}
$$

where we notice the first line corresponds to the fourth order Bartlett identity; therefore, we have $\frac{1}{T}\sum_{i=1}^{B_N}(\text{RHS of (C.71)}) = o_p(1)$. To deal with the other terms of Z_{il}, we will use the following reasoning. Let $b(l_t^{(1)},\dots,l_t^{(k)})$ be a polynomial describing a Bartlett identity (e.g., $(l_t^{(1)2}+l_t^{(2)})$). Then, provided the moments exist,

$$
E\Big(b\big(l_t^{(1)},\dots,l_t^{(k)}\big)\sum_{s<t} f_s\big(l_s^{(1)},\dots,l_s^{(k)}\big)\mid \mathcal{H}_{0,T}\Big)
$$

are m.d.s. and, therefore,

$$
\frac{1}{T}\sum_t E\Big(b\big(l_t^{(1)},\dots,l_t^{(k)}\big)\sum_{s<t} f_s\big(l_s^{(1)},\dots,l_s^{(k)}\big)\mid \mathcal{H}_{0,T}\Big)\to 0.
$$

Moreover, the existence of the variance follows from the geometric ergodicity of η_t. So we will rewrite $\sum_i Z_{il}$ as a sum of those terms. Notice the following relationship:

$$
\begin{aligned}
&4\sum_t l_t^{(1)3}\sum_{s\neq t} l_s^{(1)} + 4\sum_{t\neq s} l_t^{(1)} l_s^{(3)} + 12\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(1)} l_s^{(2)} \\
&\quad = 4\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(1)3} + 4\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(3)} + 12\sum_t l_t^{(1)}\sum_{s\neq t} l_s^{(1)} l_s^{(2)} \\
&\quad = 4\sum_t l_t^{(1)}\Big(\sum_{s\neq t} l_s^{(1)3} + \sum_{s\neq t} l_s^{(3)} + 3\sum_{s\neq t} l_s^{(1)} l_s^{(2)}\Big) \\
&\quad = 4\sum_t l_t^{(1)}\sum_{s\neq t}\big(l_s^{(1)3} + l_s^{(3)} + 3 l_s^{(1)} l_s^{(2)}\big) \\
&\quad = 4\sum_t l_t^{(1)}\sum_{s<t}\big(l_s^{(1)3} + l_s^{(3)} + 3 l_s^{(1)} l_s^{(2)}\big) \\
&\qquad + 4\sum_t \big(l_t^{(1)3} + l_t^{(3)} + 3 l_t^{(1)} l_t^{(2)}\big)\sum_{s<t} l_s^{(1)}.
\end{aligned}
$$

This term is negligible after taking the conditional expectation, summing over all i's, and dividing by T.

The remaining terms in Z_{il} are

$$6\sum_t l_t^{(1)2}\sum_{s\neq j\neq t} l_s^{(1)} l_j^{(1)} \tag{C.72}$$

$$+\,3\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(1)2} \tag{C.73}$$

$$+\sum_{t\neq s\neq j\neq k} l_t^{(1)} l_s^{(1)} l_j^{(1)} l_k^{(1)} \tag{C.74}$$

$$+\,6\sum_t l_t^{(1)2}\sum_{s\neq t} l_s^{(2)} \tag{C.75}$$

$$+\,6\sum_{t\neq s\neq j} l_t^{(1)} l_s^{(1)} l_j^{(2)} \tag{C.76}$$

$$+\,3\Bigl(\sum_{s\neq t} l_t^{(2)} l_s^{(2)}\Bigr). \tag{C.77}$$

Term (C.74) is equal to $24\sum_{t>s>j>k} l_t^{(1)} l_s^{(1)} l_j^{(1)} l_k^{(1)}$, which is a m.d.s.; therefore, $E[1/T\sum_{i=1}^{B_N}\sum_{t\neq s\neq j\neq k} l_t^{(1)} l_s^{(1)} l_j^{(1)} l_k^{(1)} \mid \mathcal{H}_{0,T}]$ is negligible. The sum of terms (C.72) and (C.76) gives $6\sum_t (l_t^{(1)2}+l_t^{(2)})\sum_{s\neq j\neq t} l_s^{(1)} l_j^{(1)}$; notice that

$$\begin{aligned}
&\sum_t \bigl(l_t^{(1)2}+l_t^{(2)}\bigr)\sum_{s\neq j\neq t} l_s^{(1)} l_j^{(1)}\\
&\quad = 2\sum_t \bigl(l_t^{(1)2}+l_t^{(2)}\bigr)\sum_{s<j,\neq t} l_s^{(1)} l_j^{(1)}.
\end{aligned}$$

Hence, $E[1/T\sum_i\sum_t (l_t^{(1)2}+l_t^{(2)})\sum_{s\neq j\neq t} l_s^{(1)} l_j^{(1)} | \mathcal{H}_{0,T}]$ is negligible.

The sum of terms (C.73), (C.75), and (C.77) gives

$$\begin{aligned}
&3\sum_t l_t^{(1)2}\sum_{s\neq t}\bigl(l_s^{(1)2}+l_s^{(2)}\bigr)+3\sum_t l_t^{(2)}\sum_{s\neq t}\bigl(l_s^{(1)2}+l_s^{(2)}\bigr)\\
&\quad = 3\sum_t \bigl(l_t^{(1)2}+l_t^{(2)}\bigr)\sum_{s\neq t}\bigl(l_s^{(1)2}+l_s^{(2)}\bigr)\\
&\quad = 6\sum_t \bigl(l_t^{(1)2}+l_t^{(2)}\bigr)\sum_{s<t}\bigl(l_s^{(1)2}+l_s^{(2)}\bigr),
\end{aligned}$$

which is a m.d.s.; again it is negligible after taking the expectation, summing, and rescaling. This completes the proof of Lemma C.12. *Q.E.D.*

We return to our goal, namely, proving that (C.15) converges to $-E(\mu_{2,t}^2)/2$.

Before we start, notice that (C.15) contains $M_i^{(j)}$, where $j = 1, 2, 3, 4$. But recall

$$\Gamma_i(R) = E(R_{i+1} \mid \mathcal{H}_{i,T}) - E(R_{i+1} \mid \mathcal{H}_{0,T}) - E(R_i \mid \mathcal{H}_{i-1,T}),$$

where we can subtract and add the term $E(R_i \mid \mathcal{H}_{0,T})$. We obtain

$$\begin{aligned}\Gamma_i(R) &= E(R_{i+1} \mid \mathcal{H}_{i,T}) - E(R_{i+1} \mid \mathcal{H}_{0,T}) \\ &\quad - E(R_i \mid \mathcal{H}_{i-1,T}) + E(R_i \mid \mathcal{H}_{0,T}) \\ &\quad - E(R_i \mid \mathcal{H}_{0,T}) \\ &\equiv D(R_i) - E(R_i \mid \mathcal{H}_{0,T}).\end{aligned}$$

Note that $D(R_i)$ has the following properties: $\sum_{i=1}^{B_N} D(R_i) = 0$ and $E(D(R_i) \mid \mathcal{H}_{0,T}) = 0$. Using (C.60), we can see that, for an arbitrary μ_i, we have

$$\text{(C.78)} \quad \frac{1}{T}\sum_{i=1}^{B_N} D(R_i)\mu_i = -\frac{1}{T}\sum_{i=1}^{B_N} D(R_i)\sum_{j<i}\mu_j - \frac{1}{T}\sum_{i=1}^{B_N}\mu_i\sum_{j<i} D(R_j).$$

Hence, if μ_i is a m.d.s. adapted to the $\mathcal{H}_{i,T}$, we obtain

$$\text{(C.79)} \quad \frac{1}{T}\sum_{i=1}^{B_N} E\big(D(R_i)\mu_i \mid \mathcal{H}_{0,T}\big) = o_p(1).$$

This property is used to eliminate terms in (C.15).

In the following, we use the compact notation $\Delta_i^{(j)}$ for the elements such that

$$M_i^{(j)} = L_i^{(j)} + \Delta_i^{(j)} \quad \text{for} \quad j = 1, 2, 3, 4.$$

Replacing the $M_i^{(j)}$ by their expressions and grouping the terms, we obtain

$$\begin{aligned}
&\big(M_i^{(1)}\big)^4 + L_i^{(4)} + 4M_i^{(1)}M_i^{(3)} + 6\big(M_i^{(1)}\big)^2 M_i^{(2)} + 3\big(M_i^{(2)}\big)^2 \\
\text{(C.80)} \quad &= L_i^{(4)} + \big(L_i^{(1)}\big)^4 + 4L_i^{(1)}L_i^{(3)} + 6\big(L_i^{(1)}\big)^2 L_i^{(2)} + 3\big(L_i^{(2)}\big)^2 \\
\text{(C.81)} \quad &\quad + 4\Delta_i^{(1)}\big(L_i^{(3)} + \big(L_i^{(1)}\big)^3 + 3L_i^{(1)}L_i^{(2)}\big) \\
\text{(C.82)} \quad &\quad + 6\big(\big(\Delta_i^{(1)}\big)^2 + \Delta_i^{(2)}\big)\big(\big(L_i^{(1)}\big)^2 + L_i^{(2)}\big) \\
\text{(C.83)} \quad &\quad + \big(\Delta_i^{(1)}\big)^4 + 4\Delta_i^{(1)}\Delta_i^{(3)} + 6\big(\Delta_i^{(1)}\big)^2\Delta_i^{(2)} + 3\big(\Delta_i^{(2)}\big)^2 \\
\text{(C.84)} \quad &\quad + 4L_i^{(1)}\big(\Delta_i^{(3)} + 3\Delta_i^{(1)}\Delta_i^{(2)} + \big(\Delta_i^{(1)}\big)^3\big).
\end{aligned}$$

The term in (C.80) once rescaled goes to zero by Lemma C.12. The term (C.81) once rescaled also goes to zero because $\Delta_i^{(1)} = D_i^{(1)}$ and (C.79) applies.

Now, we study (C.82). Note that

$$\begin{aligned}
\Delta_i^{(2)} &= \Delta_i\big(\big(M_i^{(1)}\big)^2 + L_i^{(2)}\big) \\
&= \Delta_i\big(\big(L_i^{(1)}\big)^2 + L_i^{(2)} + 2L_i^{(1)}D_i^{(1)} + \big(D_i^{(1)}\big)^2\big) \\
&= -E\big(\big(L_i^{(1)}\big)^2 + L_i^{(2)} + 2L_i^{(1)}D_i^{(1)} + \big(D_i^{(1)}\big)^2 \mid \mathcal{H}_{0,T}\big) + D_{2i},
\end{aligned}$$

where $D_{2i} \equiv D((L_i^{(1)})^2 + L_i^{(2)} + 2L_i^{(1)}D_i^{(1)} + (D_i^{(1)})^2)$. Moreover,

$$\frac{1}{T}\sum_{i=1}^{B_N} E\big[D_{2i}\big(\big(L_i^{(1)}\big)^2 + L_i^{(2)}\big) \mid \mathcal{H}_{0,T}\big] = o_p(1)$$

as a result of (C.79) and

$$\frac{1}{T}\sum_{i=1}^{B_N} E\big[\big(D_i^{(1)}\big)^2\big(\big(L_i^{(1)}\big)^2 + L_i^{(2)}\big) \mid \mathcal{H}_{0,T}\big] = o_p(1)$$

for the following reason. The second order and first order Bartlett conditions imply that $\mu_i = (L_i^{(1)})^2 + L_i^{(2)}$ is a m.d.s. So we have

$$\begin{aligned}
\frac{1}{T}\sum_{i=1}^{B_N}\big(D_i^{(1)}\big)^2\mu_i = \frac{1}{T}\sum_{i=1}^{B_N}\big(D_i^{(1)}\big)^2\sum_{i=1}^{B_N}\mu_i - \frac{1}{T}\sum_{i=1}^{B_N}\big(D_i^{(1)}\big)^2\sum_{j<i}\mu_j \\
- \frac{1}{T}\sum_{i=1}^{B_N}\mu_i\sum_{j<i}\big(D_j^{(1)}\big)^2.
\end{aligned}$$

Taking conditional expectation with respect to $\mathcal{H}_{0,T}$, the second and third terms of the sum are negligible. The sums are over different blocks and hence, due to our assumption about exponential mixing, the covariance between $(D_i^{(1)})^2$ and μ_j converges to zero sufficiently fast. We examine more carefully the first term:

$$\frac{1}{T}\sum_{i=1}^{B_N}\big(D_i^{(1)}\big)^2\sum_{i=1}^{B_N}\mu_i = \frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\big(D_i^{(1)}\big)^2\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\mu_i.$$

As μ_i is a m.d.s. with finite variance, we have

$$E\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\mu_i\right)^2 = O_p(1)$$

and

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\big(D_i^{(1)}\big)^2$$

$$=\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\left(\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}E(\eta_t\mid\mathcal{H}_{i,T})-\sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}E(\eta_s\mid\mathcal{H}_{i-1,T})\right)^2$$

$$=\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\left(\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}E(\eta_t\mid\mathcal{H}_{i,T})\right)^2 \tag{C.85}$$

$$+\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\left(\sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}E(\eta_s\mid\mathcal{H}_{i-1,T})\right)^2 \tag{C.86}$$

$$-\frac{2}{\sqrt{T}}\sum_{i=1}^{B_N}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}E(\eta_t\mid\mathcal{H}_{i,T})\sum_{s=T_{i-1}+1}^{T_i} l_s^{(1)}E(\eta_s\mid\mathcal{H}_{i-1,T}). \tag{C.87}$$

The three terms (C.85), (C.86), and (C.87) can be treated in the same manner. We will examine only (C.85):

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\left(\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)}E(\eta_t\mid\mathcal{H}_{i,T})\right)^2$$

$$=\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)2}E(\eta_t\mid\mathcal{H}_{i,T})^2$$

$$+\frac{2}{\sqrt{T}}\sum_{i=1}^{B_N}\sum_{t=T_i+1}^{T_{i+1}}\sum_{s<t} l_t^{(1)}l_s^{(1)}E(\eta_t\mid\mathcal{H}_{i,T})E(\eta_s\mid\mathcal{H}_{i,T}).$$

Note that

$$\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)2}E(\eta_t\mid\mathcal{H}_{i,T})^2\le\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)2}\lambda^{2(t-T_i)}g(\eta_t)^2\le C,$$

where g is some bounded function of η_t by the geometric ergodicity of η_t and C is a constant. Hence,

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}\sum_{t=T_i+1}^{T_{i+1}} l_t^{(1)2}E(\eta_t\mid\mathcal{H}_{i,T})^2\le\frac{1}{\sqrt{T}}\sum_{i=1}^{B_N}C=C\frac{B_N}{\sqrt{T}}=o_p(1).$$

It follows that

$$\frac{1}{T}E\left|\sum_{i=1}^{B_N}\left(D_i^{(1)}\right)^2\sum_{i=1}^{B_N}\mu_i\right| = o_p(1),$$

and

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{B_N}6E\left[\left(\left(\Delta_i^{(1)}\right)^2+\Delta_i^{(2)}\right)\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}\right)\mid\mathcal{H}_{0,T}\right]\\
&\quad=-\frac{6}{T}\sum_{i=1}^{B_N}E\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}+2L_i^{(1)}D_i^{(1)}+\left(D_i^{(1)}\right)^2\mid\mathcal{H}_{0,T}\right)\\
&\qquad\times E\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\right)\\
&\qquad+o_p(1)
\end{aligned}$$

$$=-\frac{6}{T}\sum_{i=1}^{B_N}E\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\right)^2 \tag{C.88}$$

$$\begin{aligned}
&-\frac{6}{T}\sum_{i=1}^{B_N}E\left(2L_i^{(1)}D_i^{(1)}+\left(D_i^{(1)}\right)^2\mid\mathcal{H}_{0,T}\right)E\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\right) \qquad \text{(C.89)}\\
&+o_p(1).
\end{aligned}$$

Now, we turn our attention to (C.83).
Using (C.78), we can show

$$\frac{1}{T}\sum_{i=1}^{B_N}E\left(\Delta_i^{(1)}\Delta_i^{(3)}\mid\mathcal{H}_{0,T}\right)=\frac{1}{T}\sum_{i=1}^{B_N}E\left(D_i^{(1)}\Delta_i^{(3)}\mid\mathcal{H}_{0,T}\right)=o_p(1).$$

By the geometric ergodicity of ϑ_t, we have

$$\frac{1}{T}\sum_{i=1}^{B_N}E\left(\Delta_i^{(1)4}\mid\mathcal{H}_{0,T}\right)=o_p(1).$$

The remaining terms of (C.83) are

$$\begin{aligned}
&6\left(\Delta_i^{(1)}\right)^2\Delta_i^{(2)}+3\left(\Delta_i^{(2)}\right)^2\\
&\quad=6\left(D_i^{(1)}\right)^2\left[-E\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}+2L_i^{(1)}D_i^{(1)}+\left(D_i^{(1)}\right)^2\mid\mathcal{H}_{0,T}\right)+D_{2i}\right]\\
&\qquad+3\left[-E\left(\left(L_i^{(1)}\right)^2+L_i^{(2)}+2L_i^{(1)}D_i^{(1)}+\left(D_i^{(1)}\right)^2\mid\mathcal{H}_{0,T}\right)+D_{2i}\right]^2.
\end{aligned}$$

Rearranging the terms yields

$$6\big(\Delta_i^{(1)}\big)^2\Delta_i^{(2)}+3\big(\Delta_i^{(2)}\big)^2$$

$$=3E\big(\big(L_i^{(1)}\big)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\big)^2 \tag{C.90}$$

$$+3E\big(2L_i^{(1)}D_i^{(1)}+\big(D_i^{(1)}\big)^2\mid\mathcal{H}_{0,T}\big)^2 \tag{C.91}$$

$$+6E\big(2L_i^{(1)}D_i^{(1)}+\big(D_i^{(1)}\big)^2\mid\mathcal{H}_{0,T}\big)E\big(\big(L_i^{(1)}\big)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\big) \tag{C.92}$$

$$-6\big(D_i^{(1)2}+D_{2i}\big)E\big(\big(L_i^{(1)}\big)^2+L_i^{(2)}+2L_i^{(1)}D_i^{(1)}+\big(D_i^{(1)}\big)^2\mid\mathcal{H}_{0,T}\big) \tag{C.93}$$

$$+3D_{2i}^2+6D_i^{(1)2}D_{2i}. \tag{C.94}$$

After rescaling and summing, the terms (C.91), (C.93), and (C.94) are $o_p(1)$. The term (C.92) simplifies with (C.89). Hence, the terms (C.82) and (C.83) simplify to give

$$-3\frac{1}{T}\sum_{i=1}^{B_N}E\big(\big(L_i^{(1)}\big)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\big)^2+o_p(1).$$

Now consider (C.84):

$$4L_i^{(1)}\big(\Delta_i^{(3)}+3\Delta_i^{(1)}\Delta_i^{(2)}+\big(\Delta_i^{(1)}\big)^3\big)$$

$$\begin{aligned}=&-4L_i^{(1)}E\big(L_i^{(3)}+3M_i^{(1)}M_i^{(2)}+\big(M_i^{(1)}\big)^3\mid\mathcal{H}_{0,T}\big)\\&+4L_i^{(1)}D\big(L_i^{(3)}+3M_i^{(1)}M_i^{(2)}+\big(M_i^{(1)}\big)^3\big)\\&+4L_i^{(1)}\big(D_i^{(1)}\big)^3\\&-12L_i^{(1)}D_i^{(1)}E\big(\big(M_i^{(1)}\big)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\big)\\&+12L_i^{(1)}D_i^{(1)}D\big(\big(M_i^{(1)}\big)^2+L_i^{(2)}\big).\end{aligned} \tag{C.95}$$

Note that $E((\text{C.95})\mid\mathcal{H}_{0,T})=0$ because $E(L_i^{(1)}\mid\mathcal{H}_{0,T})=0$. We obtain

$$\frac{1}{T}\sum_{i=1}^{B_N}E\big((\text{C.84})\mid\mathcal{H}_{0,T}\big)=o_p(1).$$

In conclusion, we have

$$(\text{C.15})=-\frac{1}{8T}\sum_{i=1}^{B_N}E\big(\big(L_i^{(1)}\big)^2+L_i^{(2)}\mid\mathcal{H}_{0,T}\big)^2+o_p(1).$$

Finally, note that

$$-\frac{1}{8T}\sum_{i=1}^{B_N} E\big((L_i^{(1)})^2 + L_i^{(2)} \mid \mathcal{H}_{0,T}\big)^2 \xrightarrow{p} -\frac{1}{2}E(\mu_{2,t}^2).$$

This completes the proof of Theorem 4.1. *Q.E.D.*

APPENDIX D: Optimality and Power

In this section, we prove all the results of Sections 4 and 5, except for Theorem 4.1, which has already been proved. In Section D.4, we investigate the power of our test in autoregressive models with switching mean.

D.1. *Contiguity—Proof of Corollary 4.2*

Define sequences θ_T such that

(D.1) $\quad \mathcal{N} \ni \theta_T \to \theta_0 \in \mathcal{N}.$

We establish the following intermediate result.

Corollary D.1: *For every sequence θ_T satisfying* (D.1) *and any β, the $P_{\theta_T,\beta}$ is contiguous with respect to P_{θ_T}.*

Proof: By Le Cam's first lemma (see Lemma 6.4 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 4.1, we have

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

under P_{θ_T}. It follows from (2.5) that $\mu_{2,t}(\beta,\theta_0)$ is a stationary and ergodic martingale difference sequence; hence the central limit theorem applies. We have

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

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

$$\begin{aligned} E[U] &= \exp\left(\frac{c(\beta,\beta)}{2}\right)\exp\left(-\frac{c(\beta,\beta)}{2}\right) \\ &= 1. \end{aligned}$$

Q.E.D.

PROOF OF COROLLARY 4.2: Denote

$$\ell_T(\theta_0 - d/\sqrt{T}) \equiv \frac{dP_{\theta_0 - d/\sqrt{T}}}{dP_{\theta_0}} = \frac{\prod_{t=1}^{T} f_t(\theta_0 - d/\sqrt{T})}{\prod_{t=1}^{T} f_t(\theta_0)}$$

$$= \exp\left\{\sum_{t=1}^{T}\bigl(l_t(\theta_0 - d/\sqrt{T}) - l_t(\theta_0)\bigr)\right\}.$$

Using a second order Taylor expansion around $\theta_0 - \frac{d}{\sqrt{T}}$, we obtain the following result:

For all $\theta_0 \in \mathcal{N}$, and for all vectors d,

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

uniformly (in d on all compacts) in probability.

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 mean zero and variance $E[(d' l_t^{(1)}(\theta_0))^2]$; hence we can conclude that $P_{\theta_0 - d/\sqrt{T}}$ is contiguous with respect to P_{θ_0}. Since contiguity is a transitive relationship, we may conclude from Corollary D.1 that, for all vectors d, $P_{\theta_0 - d/\sqrt{T}, \beta}$ is contiguous with respect to P_{θ_0}. This concludes the proof of Corollary 4.2. *Q.E.D.*

D.2. *Definitions and Preliminary Results on Optimality*

Denote

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

$$= \exp\left\{\sum_{t=1}^{T}\bigl(l_t(\theta_0 - d/\sqrt{T}) - l_t(\theta_0)\bigr)\right\}.$$

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

LEMMA D.2: *For all* $\theta_0 \in \mathcal{N}$, *and for all vectors* d,

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

uniformly (*in* d *on all compacts*) *in probability*.

Again, our regularity conditions guarantee the convergence of $(1/\sqrt{T}) \times \sum_{t=1}^{T} d' l_t^{(1)}(\theta_0)$ to a normal distribution with mean zero and variance $E[(d' l_t^{(1)}(\theta_0))^2]$; hence we can conclude that $P_{\theta_0 - d/\sqrt{T}}$ are contiguous with respect to P_{θ_0}. Since contiguity is a transitive relationship, we may conclude that, for all vectors d, $P_{\theta_0 - d/\sqrt{T}, \beta}$ 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

$$\theta_T = \theta_0 - \frac{d}{\sqrt{T}}, \tag{D.2}$$

$$\frac{dP_{\theta_T,\beta}}{dP_{\theta_0}} \Big/ \exp\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mu_{2,t}(\beta, \theta_T) - \frac{1}{2}E_{\theta_0}\left(\mu_{2,t}(\beta, \theta_T)^2\right) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} d' l_t^{(1)}(\theta_T) + \frac{1}{2}E_{\theta_0}\left(\left(d' l_t^{(1)}(\theta_T)\right)^2\right)\right) \to 1,$$

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

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

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

where θ_T is given by (D.2). We now want to compare tests with respect to their powers against these alternatives. In particular, we want to characterize tests by optimality properties. We start with a sequence of tests ψ_T and then show that there does not exist another sequence of tests φ_T that is asymptotically "better" for the null and all the contiguous alternatives. So let us formally define "better" tests.

DEFINITION D.3: A sequence φ_T of tests is asymptotically better than ψ_T at θ_0 if it is "better" under the null

$$\limsup \int \varphi_T \, dP_{\theta_0} \leq \liminf \int \psi_T \, dP_{\theta_0}$$

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

$$\liminf \int \varphi_T \, dP_{\theta_T,\beta} \geq \limsup \int \psi_T \, dP_{\theta_T,\beta}.$$

This definition is essentially the same as that 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 turns out to be 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 D.4: A test ψ_T is said to be admissible if there exists no asymptotically better test.

Let φ_T be some test statistic 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 $B \times \mathbf{R}^p$. Without limitation of generality, we can assume that $\nu(B \times \mathbf{R}^p) = 1$. Then let us define the loss function as

$$L(\varphi_T) = K \int \varphi_T \, dP_{\theta_0} - \int \left(\int \varphi_T \, dP_{\theta_0 - d/\sqrt{T},\beta} \right) d\nu(\beta, d). \tag{D.3}$$

By Fubini's theorem, we have

$$\begin{aligned} L(\varphi_T) &= \int \left(K - \frac{dP_{\theta_0 - d/\sqrt{T},\beta}}{dP_{\theta_0}} \right) \varphi_T \, dP_{\theta_0} \, d\nu(\beta, d) \\ &= \int \left(K - \left\{ \int \frac{dP_{\theta_0 - d/\sqrt{T},\beta}}{dP_{\theta_0}} \, d\nu(\beta, d) \right\} \right) \varphi_T \, dP_{\theta_0}. \end{aligned} \tag{D.4}$$

It follows from (D.4) that, for fixed K, $L(\varphi_T)$ is minimized by the tests ψ_T, which satisfy

$$\psi_T = \left\{ \begin{array}{l} 1 \text{ if } \left\{ \int \dfrac{dP_{\theta_0 - d/\sqrt{T},\beta}}{dP_{\theta_0}} \, d\nu(\beta, d) \right\} > K \\ 0 \text{ if } \left\{ \int \dfrac{dP_{\theta_0 - d/\sqrt{T},\beta}}{dP_{\theta_0}} \, d\nu(\beta, d) \right\} < K \end{array} \right\}. \tag{D.5}$$

So the minimal loss only depends on the *distributions* of the

$$\left\{\int \frac{dP_{\theta_0-d/\sqrt{T},\beta}}{dP_{\theta_0}}\,d\nu(\beta,d)\right\}.$$

Moreover, 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

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

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 (D.5) does not uniquely determine a test. Indeed, the behavior of the test on the event $[\{\int \frac{dP_{\theta_0-d/\sqrt{T},\beta}}{dP_{\theta_0}}\,d\nu(\beta,d)\}=K]$ does not matter. Hence the following definition will be useful.

DEFINITION D.5: The tests φ_T and φ'_T are asymptotically equivalent (with respect to the loss function L) if and only if, for all $\varepsilon>0$,

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

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 D.6: *Let ψ_T be defined by* (D.5) *and φ_T be an arbitrary test.*

(i) *If φ_T and ψ_T are asymptotically equivalent in the sense of Definition* D.5, *then*

$$\lim\big(L(\psi_T)-L(\varphi_T)\big)=0. \tag{D.6}$$

(ii) *If φ_T and ψ_T are not asymptotically equivalent, then*

$$\liminf\big(L(\psi_T)-L(\varphi_T)\big)<0. \tag{D.7}$$

Hence (D.6) *implies that ψ_T and φ_T are asymptotically equivalent.*

We conclude from Theorem D.6 that the tests ψ_T and all asymptotically equivalent sequences of tests are admissible. Any tests with genuinely better power functions would have smaller loss, which is impossible. Hence, we have to show that our test is asymptotically equivalent to tests ψ_T.

PROOF OF THEOREM D.6: It follows from (D.4) that

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

The construction of ψ_T and the fact that $0\le\varphi_T\le 1$ imply that the integrand is nonpositive. Let $\varepsilon>0$ be arbitrary. Let us define

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

Then

$$\text{(D.8)}\qquad L(\psi_T)-L(\varphi_T) = \int rI\big[|r|>\varepsilon\big](\psi_T-\varphi_T)\,dP_{\theta_0} + \int rI\big[|r|\le\varepsilon\big](\psi_T-\varphi_T)\,dP_{\theta_0}.$$

Since $|\psi_T-\varphi_T|\le 1$, we have

$$\text{(D.9)}\qquad \left|\int rI\big[|r|\le\varepsilon\big](\psi_T-\varphi_T)\,dP_{\theta_0}\right|\le\varepsilon.$$

The construction of ψ_T guarantees that $r(\psi_T-\varphi_T)\le 0$. Hence, for asymptotically equivalent tests, we have

$$\int rI\big[|r|>\varepsilon\big](\psi_T-\varphi_T)\,dP_{\theta_0} = -\int |r|I\big[|r|>\varepsilon\big]|\psi_T-\varphi_T|\,dP_{\theta_0} < -\varepsilon\int I\big[|r|>\varepsilon\big]|\psi_T-\varphi_T|\,dP_{\theta_0}\to 0.$$

This proves (D.6). For (D.7), observe that if φ_T and ψ_T are not asymptotically equivalent, then there exists an $\eta>0$ so that

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

As $r(\psi_T-\varphi_T)\le 0$, we have $rI[|r|>\varepsilon](\psi_T-\varphi_T)\le rI[|r|>\eta](\psi_T-\varphi_T) = -|\varphi_T-\psi_T||r|I[|r|>\eta]$ if $\eta\ge\varepsilon$; hence, for all ε small enough,

$$\liminf \int rI\big[|r|>\varepsilon\big](\psi_T-\varphi_T)\,dP_{\theta_0} < -\eta\limsup E_{\theta_0}|\varphi_T-\psi_T|$$

$$\times I\left[\left|\left\{\int \frac{dP_{\theta_0-d/\sqrt{T},\beta}}{dP_{\theta_0}}\,d\nu(\beta,d)\right\}-K\right|>\eta\right]$$
$$<0,$$

and together with (D.8) and (D.9), this proves Theorem D.6. *Q.E.D.*

Assume ν to be a probability measure on the product space of β and $\mathbf{R}^p$.

DEFINITION D.7: Define ϕ_T as the tests that reject when $\int \exp(Z_T(\beta,\theta_T))\,d\nu(\beta,d)>K$ and accept when $\int \exp(Z_T(\beta,\theta_T))\,d\nu(\beta,d)<K$.

We now want to show that the tests ψ_T and ϕ_T are asymptotically equivalent. A sufficient condition for asymptotic equivalence is

$$\int \exp\bigl(Z_T(\beta,\theta_T)\bigr)\,d\nu(\beta,d)\Big/\int \frac{dP_{\theta_0-d/\sqrt{T},\beta}}{dP_{\theta_0}}\,d\nu(\beta,d)\to 1. \tag{D.10}$$

We know that, for all finite sets β_i, d_i,

$$\exp\bigl(Z_T(\beta_i,\theta_0-d_i/\sqrt{T})\bigr)\Big/\frac{dP_{\theta_0-d_i/\sqrt{T},\beta_i}}{dP_{\theta_0}}\to 1. \tag{D.11}$$

So suppose that, for all $\varepsilon>0$ and $\eta>0$, we could find a partition $S_1,\dots,S_K$ so that with probability greater than $1-\varepsilon$ for all i, $(\beta,d),(\gamma,e)\in S_i$, $|Z_T(\beta,\theta_0-d/\sqrt{T})-Z_T(\gamma,\theta_0-e/\sqrt{T})|<\eta$, $|\frac{dP_{\theta_0-d/\sqrt{T},\beta}}{dP_{\theta_0}}-\frac{dP_{\theta_0-e/\sqrt{T},\gamma}}{dP_{\theta_0}}|<\eta$. Then, (D.10) will be an easy consequence of (D.11).

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.

Then, the tests ϕ_T are asymptotically equivalent to the tests ψ_T. Consequently, we have the following result.

THEOREM D.8: *Let φ_T be a sequence of tests that is asymptotically better* (*in the sense of Definition* D.3) *than ϕ_T. Then φ_T is asymptotically equivalent to ϕ_T.*

PROOF: Theorem D.6 shows that, if the ϕ_T are equivalent to the ψ_T, then

$$\lim\bigl(L(\phi_T)-L(\psi_T)\bigr)=0. \tag{D.12}$$

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

$$\liminf\bigl(L(\varphi_T)-L(\phi_T)\bigr)\geq 0. \tag{D.13}$$

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 $\int \liminf h_n \, d\delta \leq \liminf \int h_n \, d\delta$. Definition D.3 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

$$\begin{aligned} &L(\varphi_T) - L(\phi_T) \\ &\quad = K\left(\int \varphi_T \, dP_{\theta_0} - \int \phi_T \, dP_{\theta_0}\right) \\ &\qquad - \int \left(\left(\int \varphi_T \, dP_{\theta_0 - d/\sqrt{T},\beta}\right) - \left(\int \phi_T \, dP_{\theta_0 - d/\sqrt{T},\beta}\right)\right) d\nu(\beta, d), \end{aligned}$$

we can conclude that

$$\limsup\big(L(\varphi_T) - L(\phi_T)\big) \leq 0. \tag{D.14}$$

Equations (D.13) and (D.14) allow us to conclude that $\lim(L(\varphi_T) - L(\phi_T)) = 0$; hence, (D.12) also implies that $\lim(L(\varphi_T) - L(\psi_T)) = 0$. Then Theorem D.6 implies that φ_T and ψ_T are asymptotically equivalent. Since we did show that the ϕ_T are equivalent to the ψ_T, this proves Theorem D.8. *Q.E.D.*

D.3. *Proofs of Theorems 4.3, 5.1, and 5.2*

PROOF OF THEOREM 4.3: 1. Proof of (4.4):

We have to analyze the difference between $Z_T(\beta, \theta_T)$, where $\theta_T = \theta - d/\sqrt{T}$, and

$$\mathrm{TS}_T(\beta, \widehat{\theta}) = \frac{1}{\sqrt{T}} \sum \mu_{2,t}(\beta, \widehat{\theta}) - \frac{1}{2T} \widehat{\varepsilon}(\beta)' \widehat{\varepsilon}(\beta),$$

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

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.

Terms $-\frac{1}{2}E(\mu_{2,t}(\beta,\theta_T)^2)+\frac{1}{2}E((d'l_t^{(1)}(\theta_T))^2)$ are continuous functions of θ, converging uniformly in β to

$$-\frac{1}{2}E\big(\mu_{2,t}(\beta,\theta_0)^2\big)+\frac{1}{2}E\big(\big(d'l_t^{(1)}(\theta_0)\big)^2\big).$$

By Point (ii) of Theorem 3.1,

$$\frac{1}{2T}\widehat{\varepsilon(\beta)}'\widehat{\varepsilon(\beta)}\xrightarrow{P}\frac{1}{2}E\big(\mu_{2,t}(\beta,\theta_0)^2\big)-\frac{1}{2}d'I(\theta_0)d.$$

Hence it is sufficient for us to show that

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

converges (uniformly in β) to 0. Observe that our conditions guarantee that the ML estimator is $\sqrt{T}$ consistent. Hence it is sufficient to show that, for all M,

$$\sup_{\beta,\|\theta-\theta_0\|\le M/\sqrt{T}}\big|\nu_T(\beta,\theta)-\nu_T(\beta,\theta_0)\big|\to 0, \tag{D.15}$$

where ν_T was defined in (C.1). Equation (D.15) can be shown using a proof similar to that of Point (i) of Theorem 3.1.

2. Validity of the asymptotic critical values:

The validity of asymptotic critical values obtained by plugging in $\hat{\theta}$ instead of θ can be established using Theorem C.2 and the same proof as for Theorem 3.2.

3. Admissibility:

From

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

and Theorem 4.1, we can conclude that

$$\left(\frac{dP_{\theta_T,\beta}}{dP_{\theta_0}}\right)\Big/\exp\big(Z_T(\beta,\theta_T)\big)\to 1,$$

where the convergence is—again—uniform in probability with respect to P_{θ_0}. Then, the admissibility of expTS follows from Theorem D.6.

We now have shown Theorem 4.3. *Q.E.D.*

PROOF OF THEOREM 5.1: Using the same argument as in the proof of Theorem 3.1, we can show that the joint distribution of $(\mathrm{TS}_T(\beta, \hat{\theta}), \log \frac{dP_{\theta_T,\beta}}{dP_{\theta_0}})'$ converges under P_{θ_0} to a normal distribution with mean $(-\frac{1}{2}k(\beta,\beta), 0)$ and covariance

$$\begin{pmatrix} k(\beta,\beta) & k(\beta,\beta) \\ k(\beta,\beta) & k(\beta,\beta) \end{pmatrix}.$$

It follows from Le Cam's third lemma (van der Vaart (1998)) that $\mathrm{TS}_T(\beta, \hat{\theta})$ converges in distribution under $P_{\theta_T,\beta}$ to a normal distribution with mean $k(\beta,\beta) - \frac{1}{2}k(\beta,\beta) = \frac{1}{2}k(\beta,\beta)$ and variance $k(\beta,\beta)$. The same can be shown for any sequence of $\beta_i : 1 \le i \le N$, yielding the desired result. *Q.E.D.*

PROOF OF THEOREM 5.2: First of all, let us observe that

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

Let us assume that, for one h, there exist infinitely many values of ρ, so that (5.1) is fulfilled. We see that $\mu_{2,t}(\beta,\theta)$, and hence d, are analytic functions of ρ. Therefore, $E_{\theta_0}((\mu_{2,t}(\beta,\theta_0) - d' l_t^{(1)}(\theta_0))^2)$ 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. Evaluating the derivative of order $t-s$ at $\rho = 0$ yields

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

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

D.4. *Power of ExpTS Test for Autoregressive Models*

In Section 5, we established that, under certain circumstances, even local alternatives shrinking with $T^{-1/4}$ cannot be detected. Here we give a detailed description of this case when the model under the null is an autoregressive process with unknown mean. Consider the model

$$\text{(D.16)} \quad \begin{aligned} &y_t = \mu_t + u_t, \\ &\varphi(L)u_t = e_t,\ e_t \sim \text{i.i.d. } \mathcal{N}(0,\sigma^2), \end{aligned}$$

with $\varphi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^r$. We assume that the covariance of η_t is such that

$$\text{(D.17)} \quad \operatorname{Cov}(\eta_t, \eta_s) = Const\,\rho^{|t-s|}.$$

LEMMA D.9: *Consider the model* (D.16) *and* (D.17); *expTS test for testing* $H_0: \mu_t = \mu$ *has power under the local alternatives* $H_{1T}: \mu_t = \mu + \eta_t/T^{1/4}$ *if and only if* $1/\rho$ *is not a root of the characteristic equation of the AR model*, *that is*,

$$\phi_r(1/\rho)^r + \phi_{r-1}(1/\rho)^{r-1} + \phi_{r-2}(1/\rho)^{r-2} + \cdots + \phi_1(1/\rho) - 1 \neq 0.$$

We see that if the roots of the characteristic equation are all complex, then this condition is necessarily satisfied. On the other hand, if $1/\rho$ is a root of the characteristic equation, our test does not have power against alternative of order $1/T^{1/4}$. Moreover, it is impossible to construct a test that would have such power.

PROOF OF LEMMA D.9: First of all, let us define the notation *Const* as a constant (i.e., nonrandom) function of the parameters, not necessarily always the same. So when *Const* occurs twice in a formula, it does not necessarily denote the same thing.

Then we have the first order derivatives

$$\text{(D.18)} \quad \begin{aligned} &\frac{\partial l_t}{\partial \mu} = Const \cdot e_t, \\ &\frac{\partial l_t}{\partial \varphi_i} = \frac{1}{\sigma^2} e_t (y_{t-i} - \mu), \quad i = 1, \ldots, r, \\ &\frac{\partial l_t}{\partial (\sigma^2)} = Const\left(e_t^2 - \sigma^2\right), \end{aligned}$$

and the second order derivatives

$$\frac{\partial^2 l_t}{\partial \mu^2} = Const\left(e_t^2 - \sigma^2\right)$$

and $\mu_{2,t}$ as

$$\mu_{2,t} = Const\left(e_t^2 - \sigma^2\right) + Const \sum_{i \geq 1} \rho_t^i e_t e_{t-i}. \tag{D.19}$$

The problematic points for our test statistic are these ρ, where $\mu_{2,t}$ is a linear combination of the first order derivatives. So let us assume we have a ρ for which this is the case:

$$\mu_{2,t} = A_1 \frac{\partial l_t}{\partial \mu} + A_2 \frac{\partial l_t}{\partial (\sigma^2)} + \sum_{i=1}^{r} B_i \frac{\partial l_t}{\partial \varphi_i}. \tag{D.20}$$

As e_t is normal and independent from the past, it is uncorrelated with $\mu_{2,t}$, $\frac{\partial l_t}{\partial \varphi_i}$, $\frac{\partial l_t}{\partial \sigma^2}$ and we can conclude that $A_1 = 0$. Moreover, for the same reasons, $\frac{\partial l_t}{\partial \sigma^2}$ is uncorrelated with $\frac{\partial l_t}{\partial \varphi_i}$ and $e_t e_{t-i}$. Now it follows from (D.19) that

$$\mu_{2,t} - Const \frac{\partial l_t}{\partial (\sigma^2)} = Const \sum_{i \geq 1} \rho^i e_t e_{t-i}, \tag{D.21}$$

and from (D.20) (as $A_1 = 0$)

$$\mu_{2,t} - Const \frac{\partial l_t}{\partial (\sigma^2)} = \sum_{i=1}^{r} B_i \frac{\partial l_t}{\partial \varphi_i}. \tag{D.22}$$

Taking the difference between (D.21) and (D.22) gives

$$Const \frac{\partial l_t}{\partial (\sigma^2)} = Const \sum_{i \geq 1} \rho^i e_t e_{t-i} - \sum_{i=1}^{r} B_i \frac{\partial l_t}{\partial \varphi_i}. \tag{D.23}$$

As mentioned above, $\frac{\partial l_t}{\partial \sigma^2}$ is uncorrelated with all the terms on the right hand sides of (D.21) and (D.22). Hence, the left hand side of Equation (D.23) equals 0 and we have

$$Const \sum_{i \geq 1} \rho^i e_t e_{t-i} = \sum_{i=1}^{r} B_i \frac{\partial l_t}{\partial \varphi_i}.$$

The trivial possibility is that the *Const* on the left hand side equals 0. Then the right hand side equals zero, too, and this implies from (D.20) that $\rho = 0$, which corresponds to white noise η_t. We know that we can do nothing in this case.

So let us assume that *Const* differs from 0. Then we can divide by this constant, and use (D.18) to conclude that

$$\sum_{i\geq 1}\rho^i e_t e_{t-i} = \sum_{i=1}^{r} b_i e_t (y_{t-i}-\mu)$$

or

$$e_t \sum_{i\geq 1}\rho^i e_{t-i} = e_t \sum_{i=1}^{r} b_i (y_{t-i}-\mu), \tag{D.24}$$

where the b_i are coefficients proportional to the B_i. Now observe that e_t are independent from the past. So multiplying both sides of (D.24) with e_t and taking the conditional expectation with respect to the past yields

$$\sum_{i\geq 1}\rho^i e_{t-i} = \sum_{i=1}^{r} b_i (y_{t-i}-\mu). \tag{D.25}$$

Multiplying both sides with ρL, where L is the lag operator, and subtracting from (D.25) yields

$$\rho e_{t-1} = \sum_{i=1}^{r} b_i (y_{t-i}-\mu) - \rho\left(\sum_{i=1}^{r} b_i (y_{t-i-1}-\mu)\right),$$

or equivalently,

$$\begin{aligned}\rho e_{t-1} &= b_1(y_{t-1}-\mu) + (b_2-\rho b_1)(y_{t-2}-\mu) + \cdots \\ &\quad + (b_r-\rho b_{r-1})(y_{t-r}-\mu) - \rho b_r (y_{t-r-1}-\mu).\end{aligned}$$

From (D.16), we can conclude that

$$e_t = \varphi(L)(y_t-\mu).$$

Hence,

$$\begin{aligned}&\rho\big((y_{t-1}-\mu) - \phi_1(y_{t-2}-\mu) - \cdots \\ &\qquad - \phi_{r-1}(y_{t-r}-\mu) - \phi_r(y_{t-r-1}-\mu)\big) \\ &\quad = b_1(y_{t-1}-\mu) + (b_2-\rho b_1)(y_{t-2}-\mu) + \cdots \\ &\qquad + (b_r-\rho b_{r-1})(y_{t-r}-\mu) - \rho b_r (y_{t-r-1}-\mu).\end{aligned}$$

Since the $y_{t-i} - \mu$ are linearly independent, the coefficients must be equal. Therefore

$$\rho = b_1,$$

$$-\rho\phi_{i-1} = b_i - \rho b_{i-1}, \quad i = 2, \dots, r, \tag{D.26}$$

$$-\rho\phi_r = -\rho b_r. \tag{D.27}$$

Then (D.26) allows us to solve for b_i, $i = 2, \dots, r$,

$$b_i = -\rho\phi_{i-1} + \rho b_{i-1},$$

so

$$b_i = -\rho\phi_{i-1} - \rho^2\phi_{i-2} - \cdots - \rho^{i-1}\phi_1 + \rho^i.$$

This holds for all $i \le r$. So we have

$$b_r = -\rho\phi_{r-1} - \rho^2\phi_{r-2} - \cdots - \rho^{r-1}\phi_1 + \rho^r.$$

Equation (D.27), however, requires (as we assumed $\rho \neq 0$) that $b_r = \phi_r$; hence,

$$\phi_r = -\rho\phi_{r-1} - \rho^2\phi_{r-2} - \cdots - \rho^{r-1}\phi_1 + \rho^r$$

or

$$0 = \phi_r + \rho\phi_{r-1} + \rho^2\phi_{r-2} + \cdots + \rho^{r-1}\phi_1 - \rho^r,$$

which is equivalent to

$$0 = \phi_r(1/\rho)^r + \phi_{r-1}(1/\rho)^{r-1} + \phi_{r-2}(1/\rho)^{r-2} + \cdots + \phi_1(1/\rho) - 1.$$

Q.E.D.

REFERENCES

ANDREWS, D. W. K. (1999): "Estimation When a Parameter Is on the Boundary," *Econometrica*, 67, 1341–1383. [2]

——— (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734. [2]

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414. [70]

——— (1996): "Testing for Serial Correlation Against an ARMA(1, 1) Process," *Journal of the American Statistical Association*, 91, 1331–1342. [6,8]

BARTLE, R. G. (1966): *The Elements of Integration*. New York: Wiley. [18]

BARTLETT, M. S. (1953a): "Approximate Confidence Intervals," *Biometrika*, 40, 12–19. [17]

——— (1953b): "Approximate Confidence Intervals, II. More Than One Parameter," *Biometrika*, 40, 306–317. [17]

DUEKER, M. J. (1997): "Markov Switching in GARCH Processes and Mean-Reverting Stock Market Volatility," *Journal of Business & Economic Statistics*, 15, 26–34. [1]

GRAY, S. (1996): "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62. [1]
HAAS, M., S. MITTNIK, AND M. S. PAOLELLA (2004): "A New Approach to Markov-Switching GARCH Models," *Journal of Financial Econometrics*, 2, 493–530. [1]
HAMILTON, J. D., AND R. SUSMEL (1994): "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333. [1]
HU, L., AND Y. SHIN (2008): "Optimal Testing for Markov Switching GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 12 (3), Article 3. [1]
JENSEN, S., AND A. RAHBEK (2004a): "Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case," *Econometrica*, 72, 641–646. [3]
——— (2004b): "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226. [4]
LANG, S. (1993): *Real and Functional Analysis*. New York: Springer-Verlag. [11]
POLLARD, D. (1990): *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics. [22]
STRASSER, H. (1995): *Mathematical Theory of Statistics*. New York: Springer-Verlag. [70]
VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press. [25,67,76]

Département de Sciences Economiques, University of Montreal, CP 6128, succ Centre Ville, Montréal, Quebec H3C3J7, Canada; marine.carrasco@umontreal.ca,

Dept. of Economics, Wayne State University, 656 W. Kirby St., Detroit, MI 48202, U.S.A.; lianghu@wayne.edu,

and

Dept. of Economics, Washington University in St. Louis, Campus Box 1208, St. Louis, MO 63130, U.S.A.; werner@ploberger.com.

Manuscript received June, 2009; final revision received February, 2013.